

# 人脑与AI的 数学思维



我们即将进入一个由算法主导和支配的世界，AI将在互联网、绘画、音乐、写作等全方面挑战人类的创造力和想象力。

索托伊用数学思维帮助我们理解算法，引领人们认知创造力的本质，帮助人类创造一个人与机器共存的美好未来。

# The Creativity Code

How AI Is Learning to Write,  
Paint and Think



机械工业出版社  
China Machine Press

王晓燕 陈浩 程国建 —— 译

〔英〕马库斯·杜·索托伊  
(Marcus du Sautoy) 著

# 天才 与 算法



天才与算法：人脑与AI的数学思维

The Creativity Code: How AI Is Learning to Write, Paint and Think

（英）马库斯·杜·索托伊（Marcus du Sautoy） 著

王晓燕 陈浩 程国建 译

ISBN: 978-7-111-64714-0

本书纸版由机械工业出版社于2020年出版，电子版由华章分社（北京华章图文信息有限公司，北京奥维博世图书发行有限公司）在中华人民共和国境内（不包括香港、澳门特别行政区及台湾地区）制作与发行。

版权所有，侵权必究

客服热线：+ 86-10-68995265

客服信箱：service@bbbvip.com

官方网址：www.hzmedia.com.cn

新浪微博 @华章数媒

微信公众号 华章电子书（微信号：hzebook）

# 目录

赞誉

第1章 洛夫莱斯测试

第2章 激发创造力

创造力教得会吗

第3章 向人类宣战

非凡的游戏男孩

首战告捷

精妙绝伦

人类的反击

山头还是主峰

第4章 算法，网络时代的生活秘诀

荒岛算法

数学，幸福婚姻的秘诀

书商之战

第5章 从“自上而下”到“自下而上”

看还是不看

算法的幻觉

第6章 算法的进化

如果你喜欢.....

如何训练算法

偏见和盲点

机器之间的战争

第7章 数字绘画

什么是艺术

生物的创造力

可视化编码

分形：大自然的代码

从“亚伦”到“绘画傻瓜”

第8章 站在巨人的肩膀上

复活伦勃朗

创造力竞争

算法如何思考

算法就是一门艺术

第9章 数学的艺术

数学证明的游戏

数学的起源  
证明的起源  
第10章 数学家的望远镜  
Coq证明助手  
人脑的极限  
沃沃斯基的愿景  
第11章 音乐：声响的数学之旅  
巴赫：首位音乐程序员  
艾米：人工智能作曲家  
模拟游戏：音乐图灵测试  
“深度巴赫”：从头开始再生作曲家  
第12章 歌曲的创作公式  
普希金、诗歌和概率  
“续作者”：第一个人工智能即兴爵士作曲演奏者  
“心流机”  
量子作曲  
人为何创作音乐  
第13章 深度数学  
Mizar的数学  
数学图灵测试  
巴别数学图书馆  
数学寓言  
意料之外的故事  
数学的叙述艺术  
第14章 语言游戏  
《危险边缘》  
“华生”的工作模式  
在翻译中迷失  
机器人术语  
受阻于“中文房间”试验  
第15章 人工智能讲故事  
如何在一个月内完成一部小说  
哈利·波特和死亡波特尼克  
假如……  
伟大的“自动化”数学家  
人工智能新闻  
第16章 为什么要创造：思想的交流

## 赞誉

艺术一直被认为是人类情感和创造力的产物，是AI无法企及的高峰。本书系统而全面地呈现了在绘画、音乐、文学等领域中AI算法的研究，充满了逻辑性和实证性，展示了多个突破“图灵测试”的成果。对于AI研究者，这是一本AI艺术研究的“全球旅游指南”。对于艺术家，这本书是客观了解AI能力、利用AI技术提高创作效率和激发突破性创新的加速器。

英特尔中国研究院院长

宋继强

一本书读通AI算法的前世今生：从代数之父花拉子密到DeepMind创始人，都在用数学驱动世界前进。本书通过剖析人脑与AI各自的局限，打开数学、音乐、绘画和文学的奥秘之门。人类的天才创造力跟数学息息相关，而创造力是进化的根本动力。实际上，只有不断探索生物与机器算法的终极奥秘，才能最终打通大脑与AI的壁垒，世界也将迎来创造力指数级爆发的新纪元。

新智元创始人

杨静

这是一本精彩绝伦的好书！被誉为牛津“科学大使”的索托伊，将数学家、爵士乐手、作曲家、小说家、诗人、画家、软件工程师等人一系列创造行为的共同“原码”展现在读者面前，这一“原码”是人类创造力的源泉，也是鉴赏和领略AI无限魅力的金钥匙。这本书应成为每位关注AI的思想者的案头必备。

“信息社会50人论坛”执行主席，苇草智酷创始合伙人

段永朝

这是一本解释和探究如何激发AI创造力的书。通过对当下AI最新展示出来的与“创造力”相似的能力，本书反过来解释什么是人类创造力，以及如何理解这种创造力在人类大脑中产生的模式与结构。作者讨论的

是最新一代AI的“极限”：它能否拥有与人类相似甚至超越人类的创造力，并且帮助人类更深度地理解我们所珍视和渴望的“创造力”可以怎样生成、保护和提高。这就是当下“机器学习”在做的一些事情。作者运用数学、概率学、心理学等多重视角，去试图理解从原始数据开始慢慢学习的AI，是如何产生判断与决策的。对神经网络深度学习的研究，让我们超越“AI或受制于人类逻辑指导的代码，或产生碾压人类的高级智能”这两个极端的二元对立，开始去试验并尝试人类与AI之间相互学习、共同成长的方法与路径。

北京大学新闻与传播学院教授

吴靖

当今世界的风口浪尖非AI莫属。当AI的各种应用逐渐渗透到我们生活中的时候，人们自然会对AI的演变与发展产生好奇。如果你对AI的工作机理有兴趣，我推荐你阅读英国皇家学会和美国数学学会会士马库斯·杜·索托伊教授撰写的这本《天才与算法：人脑与AI的数学思维》。在本书中，作者完全未采用数学语言，却形象生动、通俗易懂地描绘了算法、数学思维和艺术创作之间的关系，值得读者从艺术的视角细细品味！

清华大学经济管理学院教授

肖勇波

AI可以从打砖块游戏和下围棋中发现妙招并战胜人类，AI还可以用凡·高的风格作画，用巴赫的风格作曲……书中这些进展让我亦喜亦悲。某一天，AI或许能够创造并教会我们参与更加波澜壮阔的“游戏”；AI或许不再模仿，而是成为我们这个时代的凡·高和巴赫！希望这一天来得晚一些。

电子科技大学互联网科学中心主任，教授

周涛

创造力是人类智慧最后的阵地，然而机器经过上亿次算法的迭代也可能完成创造它的人类无法理解的艺术创作。《天才与算法：人脑与AI的数学思维》带我们通向人类与机器共同创造之路。

北京师范大学系统科学学院教授，集智俱乐部、集智学园（北京）科技有限公司创始人

张江



# 第1章 洛夫莱斯测试

克劳德·德彪西（Claude Debussy）

艺术创造规则，而不是规则创造艺术。

“这台机器实在是太漂亮了！”17岁的阿达·拜伦<sup>[1]</sup>（Ada Byron）在参观查尔斯·巴贝奇<sup>[2]</sup>（Charles Babbage）发明的差分机模型时发出了这样的惊叹。

差分机是一种自动机械计算器，由大大小小的齿轮堆叠成塔状结构，每个齿轮的轮齿上标记有数字，旁边装有一个手柄。通过手柄的转动来驱动齿轮，在嘎吱嘎吱声中，差分机可以自动完成比较简单的数学计算，比如平方、立方甚至平方根……

受母亲及家庭老师的熏陶和培养，阿达从小就对数学和机械有着极其浓厚的兴趣，长期关注并深入研究巴贝奇的设计构想。数年后，已经嫁给洛夫莱斯伯爵的阿达决定与巴贝奇合作，致力于分析机的研发。分析机不同于差分机，它是一种通用的数学计算机，能力不局限于自动机械计算。阿达认为分析机应该发展成一个可用符号来表示任何事物的装置，她还预言分析机可以绘图、演奏音乐，甚至在其他科学领域也会有一些建树。此外，她认为分析机将产生一种全新的“理想化科学”，数学家将通过编程指挥机器去执行任务。她甚至预测，这台机器将能够“谱写”出属于科学家的“美妙乐章”。这些连巴贝奇都没有想过。

现在的观点普遍认为，阿达所阐述的核心思想是向创造性编码迈出的第一步，它启发着艾伦·图灵（Alan Turing）、马文·明斯基（Marvin Minsky）和唐纳德·米基（Donald Michie）等一系列先驱，并引发了席卷全球的人工智能革命。阿达对机器能力保持着谨慎的态度：“分析机能做我们命令它去做的任何工作，但不能创造；能分析，但不能预测。它的本质是帮我们实现已知的事情。”阿达认为，分析机有其局限性：输出（你所得到的）不会超过输入（你所投入的）。

多年来，这个观点已在计算机科学领域成为共识。它就像一种心理暗示，让我们觉得“智能机器取代人类”绝对不可能发生。也有人认为，想让机器拥有智能，前提是先搞清楚人类的智能。



研究人类智能，在宏观上有心理学，在微观上有分子生物学等学科，但每个方向在研究到一定阶段就会遇到瓶颈而停滞不前了。没有任何一个学科能告诉我们：智能从何而来，怎样才能产生智能。现在，即使我们已经了解了许多关于大脑的知识，但人类智能仍然像是一个“黑匣子”。对这个“黑匣子”的研究有两种方法：要么从外部观察其行为，模拟其结构，我们称此为自上而下的研究；要么猜测其结构，从外部的输入、输出验证其行为，我们称此为自下而上的研究。在计算机编码的思维模式方面，人们逐渐从自上而下的方式向自下而上的方式进行转变。究其原因，是因为自下而上的方式更有助于计算机寻找合适的“路径”去完成任务。事实上，我们根本不需要搞清楚什么是智能，以及怎样才能产生智能，就可以让算法在“数字之海”畅游。此外，算法还能像孩子一样学习成长。机器学习所创建的代码甚至能够产生惊人的洞察力和决策能力，可应用于医学图像精准识别及股票交易决策等领域。新一代的程序员用实践证明了阿达·洛夫莱斯的预言是错误的——输出可以大于输入！你可以获得比输入更多的东西。

尽管如此，我们仍然相信人类活动的某些领域是机器依然无法触及并参与的，比如人类的创造力。人类以非凡的想象力和创新力，创造出反映内心情感世界的艺术作品，进而拓展了人类存在的意义。换言之，如果把人类智慧看成一种抽象意义上的编码，那么艺术作品就是这段编码中情感的流露和表达。

这完全取决于人类的智慧。莫扎特的《安魂曲》暗含了对死亡的思考；莎士比亚的《奥赛罗》含有对复杂情感的体会；伦勃朗的肖像画所描绘的模特衣着和样貌，也蕴含着人物情感的倾诉。机器的创造力怎么有可能取代人类，或者说去和莫扎特、莎士比亚、伦勃朗这些艺术巨匠竞争呢？

需要声明的是，本书涉及的艺术主要源于西方文化，这主要是因为这些绘画、音乐、文学作品都是我所熟知的。首先，我对自己以西方为中心的观点表示歉意。其次，跨越不同文化背景目前仍存在巨大的挑战。因此，需要一个统一的标准来衡量人类与计算机的创造力。

从广义上来讲，人类的创造力不仅仅存在于艺术领域：鼎鼎有名的肥鸭餐厅（米其林三星级餐厅）的主厨赫斯顿·布卢门撒尔（Heston Blumenthal）的分子烹饪法，“荷兰飞人”“巴萨教父”约翰·克鲁伊夫（Johan Cruyff）精湛绝伦的足球技法，英国女建筑师扎哈·哈迪德（Zaha Hadid）美轮美奂的建筑设计，匈牙利发明家、雕刻家、建筑学

教授厄尔诺·鲁比克（Ernö Rubik）发明的魔方，甚至游戏《我的世界》的代码都应该被视为人类创造力的一部分。

在我所研究的数学世界里，创造也是重要的组成部分。创造的吸引力是使我伏案数小时构思方程式、写出证明过程的重要原因之一。经过漫长的酝酿、不懈的努力，灵感瞬间闪现，可以使我构思并绘制出一个新的对称结构。这个从无到有的过程，给我带来纯粹的兴奋体验，而这正是创造力的魅力所在。

创造力是一种原动力，它驱使人们产生新的、令人惊讶的、有价值的想法，并积极地将这些想法付诸实践。

创造出在表面上看似新的东西相对容易，比如我可以借助计算机得到无数种关于对称性结构的建议方案。但这是在意料之中的，不会令人惊讶，更无价值可言。若我发现的不仅是一种新的对称结构，而且在群论与数论之间建立了关联，为探索充满未知的数学世界提供了一种新的途径，那这就是它令人惊讶的价值。

每当我们自以为预见到“故事的结局”时，就会突然被带往下一个新的开始。这是一份值得我们充分关注的大惊喜。所以，无论是我还是其他任何人，在遇到偶然间的创造性行为时，都会表现得异常兴奋、急迫和匆忙。

是什么赋予了事物价值？是价格还是别人的认可？价值是相对的：我写的诗或画的画，在我看来是宝贝，在别人看来可能一文不值；一部情节曲折的小说与能改变人们生活体验的文学创作方法、建筑设计方案和音乐创作手法相比，其价值就显得微不足道了。这就是伊曼努尔·康德<sup>[3]</sup>（Immanuel Kant）所说的“典型的原创力”——人类所独有的，能成为别人灵感的原创性行为。

所有这些归根结底都是神经元活动和化学反应的产物，是人类经过数百万年进化所形成的大脑的“代码”。科学家经过长期的研究发现，这一过程是有章可循的。那么，人类创造力的“算法”或它所遵循的规则是否会超乎我们的想象？

本书主要讨论新一代人工智能（AI）的“极限”：它是否能拥有与人类相似甚至超越人类的创造力。机器同样可以画画、创作音乐或小说吗？它也许无法与毕加索、莫扎特、莎士比亚等大师相媲美，但它在创

作故事、画画方面能否达到孩子的创造力水准呢？对那些或打动人心，或平淡无奇的艺术作品进行对比分析之后，机器能否学会创作？人工智能是否能帮助人类提升创造力，从而帮助我们另辟蹊径？

创造力并不局限在艺术领域发挥作用，还广泛地在其他领域起着不可比拟的作用。比如，我女儿用乐高积木搭建的城堡；我儿子帮助他的球队赢得足球比赛后，被称为“富有创造力的中场球员”。我们在创造性地解决日常问题，创造性地管理和组织团队。正如我将要在本书中阐述的，数学是一门极具创造性的学科，它与艺术有很多共通之处。

创造的冲动是人类有别于其他动物的关键要素之一。但因为人类更习惯去做重复性的日常工作，创造的“双翼”会被这种惯性束缚而得不到激发。想要创造，我们必须走出思维定式。或许此过程中机器会发挥作用，提出一些新的想法和建议，以启发我们突破思维定式，防止我们陷入简单的重复。最终，机器可能会帮助我们，而作为人类，我们应该表现得不那么像机器。

你可能会有一点诧异：作为数学家的我为什么会带你踏上了解机器创造力的旅程？原因很简单：算法、计算机代码、人工智能、机器学习都是以数学为核心的。如果想了解掌控现代生活的算法完成这些事情的内在机理，那么就需要理解支撑它们的数学规则，否则就会在未来的生活中茫然无措。

人工智能正在向人类智能发起挑战，很多人类从事的工作，人工智能也可以做，甚至做得更好。但本书的重点不在于“无人驾驶”或“智能医疗”，而是去探索“机器编码”与“人类编码”之间的竞争：计算机有创造力吗？有创造力意味着什么？我们对艺术的情感反应中有多少属于大脑对模式和结构做出反应的产物？

这不只是一个有趣的智力问答。艺术是人类大脑编码的一种“输出”，通过它，我们可以更深入地了解复杂的人类大脑是如何工作的。计算机也是如此，我们将通过计算机“创作”的艺术作品，了解计算机的工作机理。“程序员通常并不真正理解最终的代码是如何工作的”，这是“自下而上”编码方式面对的挑战之一。计算机创造的艺术像魔法师手中转动的水晶球，透过它，我们可以分析计算机是如何做出决策，并预知新代码在潜意识下的决策方向的。此外，它还可能揭示人类尚不能完全理解的计算机代码中固有的局限性和危险性。

促使我踏上研究机器创造力旅程的另一个私人原因是，作为一个数学家，我正在经历一场“生死存亡”的危机：随着人工智能的发展，在未来几十年里，数学家的工作是否还会对人类有价值？数学是一门关于数字和逻辑的学科，可这不是计算机最为擅长的吗？

然而，数学不仅是数字和逻辑，也是一门具有高度创造性，涉及美和美学的学科。这也是我反对计算机进入数学领域的部分原因。我们在研讨会和期刊上分享的不仅仅是转动机械手柄得出的计算结果。直觉和艺术敏感性是一名优秀数学家应具有的重要特质，而这一特质永远无法编入机械计算的程序中。那么，人工智能可以吗？

作为一名数学家，我想知道新一代人工智能进入世界各地的画廊、音乐厅和出版社后，会有多么成功？伟大的德国数学家卡尔·魏尔斯特拉斯（Karl Weierstrass）曾经写道：“不具备诗人气质的数学家，永远不会成为真正的数学家。”正如阿达·洛夫莱斯完美概括的那样：你需要调和，配方中要有一些拜伦的艺术风味，也需要一些巴贝奇机械的成分。虽然她认为这些由齿轮构成的机器能做的事情是有限的，但她同时也意识到它们在表达艺术特质方面的潜质：

它们可能会应用于数字以外的其他事物……例如，音乐中的和声布局，作曲的旋律安排，根据需要进行改编。分析机可能会生成任意复杂程度、精细程度的音乐作品。

阿达坚信，任何创造性的行为都取决于程序员（人），而非机器。但新一代的程序员则不这么认为，他们觉得“代码”也能胜任创造性的工作。

在人工智能诞生之初，艾伦·图灵就提出了评价计算机智能的测试方法。这里，我想提出一个新的测试方法——洛夫莱斯测试。通过该测试的规则是，算法需要创作一件艺术作品，人类程序员无法解释该算法的工作机理，但整个过程是可复现的（以排除因硬件错误产生的结果）。我们希望机器创造新的、令人惊讶的、有价值的东西。然而，对真正具有创造力的机器，还需要额外增加一条：具备超越程序员或数据集创建者的创造力。这也是阿达·洛夫莱斯所说的“难以克服的挑战”。

[1] 英国数学家，计算机程序创始人，被称为“软件之母”。她是著名英国诗人拜伦之女，结婚后被称为阿达·洛夫莱斯（Ada Lovelace）。——译者注

[2] 英国数学家、发明家、计算机先驱。——译者注

[3] 德国作家、古典哲学创始人。——译者注

## 第2章 激发创造力

巴勃罗·毕加索（Pablo Picasso）

理智是创造力最大的敌人。

现代人非常重视创造力。许多思想家、作家都曾在著作中对创造力与其重要性，以及如何激发创造力进行过阐述。在英国皇家学会（Royal Society）一次主题为“机器学习如何影响未来”的委员会议上，我认识了认知科学家玛格丽特·博登（Margaret Boden），并了解了她的观点和理论。我认为她对机器创造力的定位和评价是最中肯、最贴切的。

博登是一位独具慧眼的科学家，她的研究涉及哲学、心理学、医学、人工智能、认知科学等领域，并能将其很好地融会贯通。80多岁高龄、满头银发的她，老当益壮、精神矍铄，依然拥有着异常活跃的思维，满怀热情地参与到关于这些“锡罐”（博登对计算机的称呼）未来前景的研究和讨论当中。博登将人类的创造力归纳为三种：探索型创造力、组合型创造力、变革型创造力。

探索型创造力是探索已知事物的外部边界，在保持规则约束的同时扩展其可能的极限。巴赫的音乐创作在巴洛克时期是登峰造极的，他的复调作品通过多旋律、不同调性的应用探索音乐的世界。巴赫的前奏曲和赋格曲在调性、和声对位技巧的运用等方面涉及了所有的可能性，突破了巴洛克音乐创作的藩篱，为以莫扎特和贝多芬为代表的、由单一旋律支配的主调音乐为主流的古典主义奠定了坚实的基础。雷诺阿

（Renoir）和毕沙罗（Pissarro）的画作，重构了我们具象化的自然及周围的世界。克劳德·莫奈（Claude Monet）彻底打破了古典主义审美（高完成度）和印象审美之间的界限——他喜欢极多层次的堆叠，画作色彩饱和，纯度极高。莫奈使用厚涂法为细小的区块着色，一遍又一遍地绘制《睡莲》<sup>[1]</sup>，直到色块融入一种新的抽象思维形式——运用纯色的、感性的笔法来分解光与色的区块，消除了古典主义画风的阴影和死板的轮廓。

数学是探索型创造力的狂欢，“有限单群分类”是它的力作之一。从对称性的简单定义（4个基本性质）开始，数学家用了150年的时间，列



出了所有可能的有限单群，并最终发现“大魔群”——其元素数量超过地球上所有原子的总数，是最大的散在单群。数学创造力在挑战极限的同时，又必须遵循游戏规则。就像一名闯入秘境的探险家，仍然不能完全摆脱地球的束缚。

博登认为，人类97%的创造行为属于探索型创造行为。计算机的运算速度远胜人类，所以用“穷举”或“暴力破解”的方法求取一组模式或规则的极限也是它最擅长的。但其结果均在情理之中，并无意料之外的惊喜，不能算真正意义上的创造。

第二种创造力是组合型创造力。将两种完全不同的结构或理念结合起来是艺术家惯用的创作手法，某领域特定的规则会与其他领域构建出新的框架。组合，也是数学领域里一个非常强大的创新工具。庞加莱（Poincaré）提出的关于探寻宇宙形状的猜想，最终也是运用不同领域的工具（微分几何学、热力学）来证明的。格里戈里·佩雷尔曼（Grigori Perelman）的创造性工作，让我们意识到液体在表面上的流动方式能够对可能存在的表面进行分类（他创造性地运用微分几何学的知识，解决了庞加莱猜想）。

我的研究就是将数论中分析素数的方法运用于探索可能的对称性分类。数字与几何对称性乍看起来毫无联系，但如果用对称元代替素数，再运用这种探索素数之谜的“语言”进行分析，就可以获得关于对称性理论令人惊讶的新发现。

这种“组合”也极大地促进了艺术的创新。菲利普·格拉斯<sup>[2]</sup>（Philip Glass）在与拉维·香卡<sup>[3]</sup>（Ravi Shankar）的合作中获得了灵感，在作品中融入了多种不同风格的音乐元素，最终形成了他独具一格的创作风格：“简约音乐”。扎哈·哈迪德在她所钟爱的俄国画家卡西米尔·马列维奇（Kasimir Malevich）的启发下，形成了独特的“曲线美”建筑设计风格，甚至烹饪界久负盛名的大厨也常常在美食的创新中融合世界各地不同的烹饪风格。

一些有趣的迹象表明，组合型创造力也可能完美适合人工智能世界。比如，蓝调音乐与布列兹（Boulez）的奇特组合可能会打开声音新世界的大门，当然，也可能会生成一种不和谐甚至刺耳的噪声。这就需要程序员找出某种有趣的方式、方法来融合两者了。

博登归纳的第三种创造力更加神秘和难以捉摸，被称为变革型创造

力。这种创造力是完全改变游戏规则的创新。每种艺术形式都有这种创造力的体现，比如毕加索的立体主义绘画、勋伯格<sup>[4]</sup>（Schoenberg）的无调性音乐、乔伊斯·卡罗尔·奥茨（Joyce Carol Oates）的现代主义小说，等等。这类似于水突然从液态转变为气态的物理学相变。就像歌德描述他创作《少年维特之烦恼》的那两年时光时所说的，他脑海中突然闪现出一个画面（某种刺激激发的偶然事件）：那一刻，维特的计划落空了；仿佛有子弹从四面八方而来，汇聚成团，就像瓶中的“过冷水”，受到轻微的震动即刻成冰。

在多数情况下，变革型创造力需要彻底打破游戏规则，或者抛开前人已做的努力。很多规则乍看起来似乎已经很难突破。比如，任何数的平方都是正数，所有分子都是长线型结构而不是链结构，音乐的创作必须建立在和声音阶的结构中，脸的两侧都长着眼睛。但激发变革型创造力的“元规则”就是打破常规，去掉约束，看看到底会出现什么。艺术这种创造性行为就是通过打破已有规则或者引入新规则，最终得到了一个全新的、有价值的作品。

“-1的平方根”的引入，是16世纪中叶数学界一个很典型的变革型创新。笛卡尔将其称为虚数（虚暗含不存在的意思），这个称谓沿用至今，可许多大数学家都不承认虚数的存在。但事实证明，虚数与已有的数学理论并不矛盾。如果给计算机预设了负数不存在平方根的规则，那它就不可能建立-1平方根的概念。真正的创造性行为有时需要打破常规，创造一个新的现实世界。这能否通过一个复杂的算法来实现呢？

浪漫主义音乐的出现许多方面都打破了巴洛克风格、古典主义音乐创作的常规。像舒伯特（Schubert）这样的新秀没有沿用古典主义作曲家推崇的不同主调之间对位与平衡的技巧，而是刻意选择了出人意料的方式进行调式的运用。舒曼（Schumann）没有采用海顿（Haydn）或莫扎特一般四平八稳的和弦进行创作。肖邦（Chopin）在他不同寻常的重音段落和扭曲的节奏里创作出密集的半音跑动，在节奏上让人们大吃一惊。从中世纪到巴洛克风格，到古典主义，再到浪漫主义；从印象派到表现主义，再到.....音乐时代的更迭，是打破原有常规的过程：后来者都要参考前者才能展现它的创造性。毋庸置疑，历史背景在我们定义新事物时起着重要的作用。创造力不是绝对的，而是相对的。我们在我们所在的文化和参照系内具有创造性。

计算机的出现能否启动“相变”，激发创造力，将我们带入全新的音乐或数学世界呢？这可能很有挑战性。算法根据数据间的交互关系自主

学习，难道就不会产生更多相同的事物吗？

正如毕加索所说：“理智是创造力最大的敌人。”从表面上看，这或许悖于机器文明。但是，通过编程创建一个“元规则”，指导机器改变路线，最终产生“非理性”的行为，这是机器学习非常擅长的。

[1] 莫奈的《睡莲》系列，1897~1926年共画了181幅。——译者注

[2] 美国作曲家。他的创作融合了摇滚乐、非洲与印度音乐、西方古典音乐的元素，作品经常重复简短的旋律和节奏模式，同时加以缓慢渐进的变奏，被称为“简约音乐”。——译

[3] 印度著名演奏家、作曲家，在把印度音乐推向西方世界的过程中起到先驱的作用。——译

[4] 20世纪初奥地利表现主义作曲家，新维也纳乐派的核心人物。表现主义音乐反对印象主义注重描绘客观事物的外在，而注重表现人的内在真实，以抽象、夸张、扭曲、怪诞的手法，用刺激、不和谐的音响来写人们心中的孤寂、恐惧、绝望甚或疯狂。——译者注

## 创造力教得会吗

许多艺术家喜欢夸大其词，将他们获得创作灵感的过程说得神乎其神。在古希腊，诗人被认为是缪斯女神的附身，是缪斯将灵感注入诗人的心灵，使诗人达到一种几近疯狂的状态。柏拉图也曾说过：“诗人是神圣的，只有当他受到神的启发而失去理智时，才能创作。没有神之伟大的指引，就没有艺术。”著名的印度数学家拉马努金（Ramanujan）将他的伟大发现归功于其所信奉的印度教的娜玛卡尔女神，认为娜玛卡尔女神在睡梦中的指引使他获得了灵感。创造力是一种“疯狂”吗？或者说它真的是神的恩赐吗？

卡尔·弗里德里希·高斯（Carl Friedrich Gauss，我心目中的数学英雄之一）最不善于表述促使他产生创造性思维的灵感来源。高斯于1801年<sup>[1]</sup>发表了其一生中最伟大的著作《算术研究》（Disquisitiones arithmeticae）——被誉为现代数论的开山之作。当人们想要通过阅读这本书来找寻他独特见解的来源时，他们备感无力。由于全书包含七部分内容，所以该书被大家风趣地称为“加了七道封印的巨著”。没有人真正知道高斯的“魔力”是如何产生的，他就像魔术师从帽子里变出一只接一只的兔子一样。受到质疑时他反驳道：“房子建成后，建筑师们也不会把脚手架丢到工地上。”像拉马努金一样，高斯把自己的成功部分归功于“上帝的恩典”。他说道：“这条线将我所知道的和使我成功的因素联系了起来，但我很难说清楚它到底是什么，也弄不明白它的本质是什么。”

大师们可能无法阐明他们的想法来自何处，但这并不意味着他们没有遵循规则。艺术是构成潜意识思维过程的无数个“逻辑门”的意识表达。当然，高斯的想法也可以通过一条逻辑线索得到解释：也许他确实不具备清楚表述自己创造思维过程的能力，也许他只是想保持神秘以维护自己创造性天才的形象。柯勒律治（Coleridge）曾表示服用鸦片酊后的睡梦联翩，是诗歌《忽必烈汗》的创作之源。可惜的是，这个梦境被一个来自波洛克的人的突然造访而打断了。<sup>[2]</sup>当然，这是一个很好的故事。但这一说法也掩盖了其创作过程中的所有前期准备——长期以来孜孜不倦的努力。即便是我自己，对创造过程的描述也只关注灵感的闪现，而不是多年的辛勤努力。

人们习惯于把富有创造力的天才传奇化。坦白地说，孤独的艺术

闭门造车是一个神话。在绝大多数情况下，创造是日积月累的过程，而非一步登天。英国音乐家布莱恩·伊诺（Brian Eno）创造了“场天才”（scenius）这个术语，矛头直指“天才”（genius）。他认为群体所构成的“场”（scene）能够激发创意，帮助创造者走向成功。美国作家乔伊斯·卡罗尔·奥茨对此表示赞同：“创造性工作就像科学工作一样，应被视为一种集体的努力——一种个体发出多种声音的尝试，一种综合、探索和分析的尝试。”

怎样才能激发创造力？编程能激发机器的创造力吗？有没有什么办法或规则可以让我们变得更加富有创造力？换句话说，创造力是一种可以后天培养的技能吗？有人会说，教授知识或编写程序都只是为了使人或机器学会模仿并按照规则做事，这与创造大相径庭。然而，我们身边有许多相反的例子：有创造力的个体通过不断地学习，最终改进了他们的技能。通过行为的研究和模仿，我们最终能成为有创造力的个体吗？

每个新学期伊始，我都会问我的学生这样的问题。数学专业的博士生必须创建一个新的数学结构才能获得学位，即他们必须做出前人未做出的事。我的职责就是教他们如何做到这一点。即使答案已了然于胸，解决问题仍需要个人的创造力。

当然，他们已然受过一定程度的训练，这些训练是进入未知世界的绝对的先决条件。通过学习他人如何实现突破，可以为自己的创造力找到一个适合的环境来扎根、发芽。但这不能保证创造力的种子就一定能萌发。有些人能在某个领域拥有极佳的创造力，而在别的领域却不行。我不可能从大街上随意选一个人带走，然后教他成为一个富有创造力的数学家。也许经过十年的训练他可以做到这一点，但并非每个大脑都能拥有数学的创造力。究竟是什么让一个人成为国际象棋冠军，而另一个人却成为诺贝尔文学奖得主，这一点很是令人费解。

博登认识到，创造力不仅仅是如何成为莎士比亚或爱因斯坦。她定义了“心理创造力”和“历史创造力”，并明确了它们之间的区别与联系。自我创造力的实现结果对个体来说可能是全新的，但纵观历史其实已算是“明日黄花”，这就是心理创造力的概念。通过个人创造力的反复实践，最终创造出一些史无前例的、有价值的东西，这就是历史创造力的体现。虽然极其罕见，但它对心理创造力起到很好的激发和鼓励作用。

遵循博登提出的三种创造力策略是我激发学生创造力的秘诀。“探索”是最显而易见的途径。首先要了解我们是如何走到今天这一步的，



然后试着进一步突破这个界限。这需要对已有的知识进行深入了解。在深入探究的过程中，可能会发现一些以前未被发现的事物。创造行为是一个循序渐进的过程，“大爆炸”这种状况很少发生。就像凡·高（Van Gogh）所说的：“伟大的成就不是靠一时冲动就能实现的，它是一系列小事的日积月累。”

博登的第二种策略——“组合”，是激发新想法的有力武器。我经常鼓励学生参加研讨会，阅读那些似乎与他们所研究的问题无关的论文。来自数学不同领域的研究策略可能与他们所研究的问题产生共鸣，从而激发出新的想法。当今，科学界一些最伟大的创造都发生于不同学科的交叉领域。越能突破自己的圈子，分享自己的想法和问题，就越有可能获得更多的创造力。这就是“最容易摘到的果子”。

表面上看，“变革”是一种很难驾驭的创造力激发策略，但我们的目标是通过去掉一些已有的限制来对现有的知识体系或架构进行测试和分析。如果改变一些已被普遍接受的学科结构的基本规则，试试看会发生什么？这种方式是极具危险性的，因为撼动系统的基石很可能会导致整个系统的崩塌。但是，这让我想到培养创造力所需的另一个最重要的因素——“接受失败”。

除非你做好了失败的准备，否则就请你不要去冒那些会让你突破自我、创造新事物的风险。教育系统和商业环境两个领域非常痛恨失败，这对创造力的激发有害无利。对我的学生来说，庆祝失败与庆祝成功同等重要。当然，失败不会被写进博士论文，但我们可以从中学到很多。与学生见面时，我经常重复地说贝克特（Beckett）的那句至理名言：“失败，再次失败，却是以更好的方式失败。”

这些策略可以编写进计算机代码吗？在过去，自上而下的编码方式意味着代码的输出结果几乎不会出现创造行为。程序员们从未对其算法产生的结果感到惊讶，因为一切都是意料之中的事情，没有试验或者失败的可能性。但最近这一切都改变了：因为一种可以“从失败中吸取经验教训”的代码出现了，它前所未有的举动震惊了它的创造者，这说明它具有不可估量的价值。该算法赢得了一场此前大多数人认为机器无法精通和掌握的比赛，因为这是一场需要创造力的“游戏”。

该代码的突破性，引发了我作为数学家的生存危机。

[1] 原文年份有误，翻译时进行了更正。——译者注



[2] 柯勒律治因身体欠佳避居乡间，一日他正在批阅萨缪尔·珀切斯的《珀切斯朝圣记》，读到书中关于忽必烈上都的文字时，身体忽感疼痛，便遵医嘱服食少量鸦片酊，很快就睡意昏沉，坠入梦乡。他在梦中见到了忽必烈的夏宫，并创作了长达两三百行的诗。醒后柯勒律治神思恍惚，但梦中的意向和所写的诗句仍历历在目，于是他奋笔疾书。可就在这时，一个来自波洛克的人因事来访，他不得被对方耽搁两个多小时。待返回自己的房间后，他又惊又愧地发现，尽管自己对灵视所得还有一个模糊朦胧的印象，可除了八九十行散落的诗句，其余的诗句已在脑海中悉数消失了。——译者注

## 第3章 向人类宣战

保罗·克利（Paul Klee）

我们在持续重构世界秩序，直觉始终被尊崇。

人们经常拿数学与国际象棋做比较，认为两者之间存在着某种联系。尽管1997年计算机“深蓝”（Deep Blue）击败了顶尖国际象棋手，但机器取代数学研究机构还言之尚早。下国际象棋与数学的形式化证明颇有相似之处，但学者认为中国围棋的思维方式更能够体现数学家思考的创造性和直觉力。

我在大学本科期间访问剑桥大学数学系时首次接触了中国围棋。当时，我正在考虑在攻读博士学位期间，是否能与完成有限单群分类（对称结构的元素周期表）的科研团队合作。当我和该项目的两位创始人约翰·康威（John Conway）、西蒙·诺顿（Simon Norton）高谈阔论、展望数学的未来发展时，邻桌的学生们正在玩一种我没见过的棋盘游戏，清脆的落子声不时将我的思路打断。

我忍不住好奇，问康威：“他们在干什么？”他告诉我：“他们在下中国围棋，这是一种古已有之的益智游戏。围棋的规则非常简单，是一种相互争夺地盘的游戏，对弈双方按黑白交替的顺序将棋子落在棋盘格线的交叉点上，想方设法用自己的棋子围困住对方的棋子，直到将其吃掉。比赛的胜负主要以吃掉对方棋子数量的多少来评定。围棋最精妙的地方就在于，当你围吃对手棋子时必须避免自身不被围吃。”

围棋很像数学，可以在相当简单的规则下形成精妙绝伦、错综复杂的推理。某一天，因缘际会，这两位数学家边喝咖啡边观看围棋比赛，突然地灵光一现，康威在棋局演变的启发下创立了新的数字系统“超现实数”。

我一直着迷于各类游戏，每次外出旅行时都喜欢学习当地流行的游戏并将其带回家。所以，从剑桥回到牛津后，我就去玩具店买了一副围棋，想要研究一下学生们沉迷于它的原因。和牛津的同学下了一段时间的围棋后，我逐渐了解到它的妙趣所在。随着棋盘上的棋子越来越多，棋局变得越来越复杂，以至于很难找到一种一眼看去就了然于胸的制胜

策略。国际象棋与围棋相比，则是随着棋子一个个被吃掉，棋局变得越来越简单。

据美国围棋协会（American Go Association）估计，围棋的可能走法数量是一个大约有300位的数字。而计算机科学家克劳德·香农（Claude Shannon）估计的国际象棋走法数量约为120位（称为香农数）。这两个数字都非常庞大，它们直观反映了两种棋类游戏所有可能的走法。

我在幼年时期经常与人下国际象棋，很喜欢推演棋步。这种思维训练逐渐激发了我的数学潜质。国际象棋的行棋步骤以一种可控、有序的方式逐级建立分支，最终形成一个包含各种可能性的树状结构，计算机甚至人类都可以根据逻辑规则逐级分析不同分支的蕴含关系。相较之下，围棋就不是一种易于推算下一步行棋对策的游戏了，我们很难建立围棋行棋可能性的树状图。围棋棋手推演下一步落子策略的过程似乎更依赖于自身的直觉判断。

人类的大脑可以敏锐地捕捉到视觉图像所呈现出的结构和模式，所以围棋棋手可以通过观察棋子布局来推断棋势，然后得出下一步的应对策略。但是，计算机实现视觉处理却是几十年来一直困扰工程师们的重大技术难题之一。

人类大脑的视觉结构处理能力作为一种基本的生存技能，经过数百万年的进化已经变得高度发达。任何动物的生存能力在一定程度上都取决于它在形态万千的自然界中对不同结构图像的识别能力：原本平静的丛林之中激起的一丝混乱，极有可能预示着另一种动物的潜入。这类敏感信息备受动物们的关注，因为它关系到自己会成为猎物还是猎食者，这就是大自然的生存法则。人类的大脑非常擅长识别模式并预测它们的发展方向，同时做出适当的反应。这是人类的宝贵财富，此外，它还关系到我们对音乐和其他艺术的鉴赏能力。

对模式的识别也恰恰是作为一名数学家的我探索“数学丛林”这片充满未知的领域的重要工具。只在局部环境中按部就班地进行逻辑分析走不了太远，必须与发现“可能存在物”的直觉相结合才有可能取得显著的突破。而这种直觉正是长久以来通过对已知领域的观察和探索而建立的。

通常，很难从逻辑上去解释你所感兴趣的领域为什么具有研究价

值。有些数学猜想虽然未得到证明，但提出猜想的数学家经常能感觉到在他的论述中暗含着某种真理。正因为如此，当我们在“丛林”中摸索前行，寻求一条新的道路时，观察和直觉是相辅相成的。

善于提出好的猜想的数学家比善于证明猜想的数学家更值得尊敬。如果把围棋棋局中赢棋的最后一步落子位置比作一种猜想，那么证明猜想的过程就是行棋的过程，在这个过程中寻求赢棋的模式是非常困难的。

因此，尽管国际象棋有助于解释数学的某些特性，但围棋游戏所蕴含的智慧与数学家们在实际工作中的思维方式更为接近。这就是为什么当“深蓝”击败人类顶级的国际象棋大师时，数学家们并不感到意外。因为，围棋才是计算机的真正挑战。

几十年来，人们一直认为计算机无法穷尽围棋里的各种变化，计算机永远都无法下好围棋。所有想要创新的程序员都挑战过这个命题，但即使是一个初级的棋手似乎也能轻松胜过最复杂的计算机算法。所以，数学家们还可以躲在围棋这块“遮羞布”后面沾沾自喜：如果计算机下不了围棋，那么就意味着它没有机会挑战更加古老和精妙的“游戏”——数学。

束缚我们认知的障壁，会在计算机技术日新月异的发展中被瞬间攻破。

## 非凡的游戏男孩

2016年初，某公司发布了一款围棋程序，并宣称该程序能与世界上顶级的人类棋手较量。鉴于过去的挑战都以失败告终，围棋爱好者们对此事均持怀疑态度。因此，该公司提出组织一场人机围棋对抗赛来证明自己所言非虚。此次比赛是公开赛，且奖金丰厚，采用五番棋赛制，人类如果获胜，将得到100万美元的奖金。来自韩国的世界冠军李世石（Lee Sedol）接受了挑战。这个计算机程序就是AlphaGo。

AlphaGo是戴密斯·哈萨比斯（Demis Hassabis）智慧的结晶。哈萨比斯1976年生于伦敦，父亲是希族塞人，母亲是新加坡人。用哈萨比斯的话说，他的父母都是教师，也都是波西米亚式<sup>[1]</sup>的技术恐惧症患者。他的弟弟和妹妹都走上了创作之路，一个成为作家，另一个成为作曲家。所以，哈萨比斯搞不清楚他的科学“极客”潜质遗传于何处。早在童年时期，哈萨比斯就显露出自己的天赋，尤其在游戏方面。11岁时，他的国际象棋排名就已经位列同龄段世界第二了。

在列支敦士登参加一场国际巡回赛时，哈萨比斯的脑海中突然闪现出这样一个念头：“在偌大的比赛大厅里，几百名来自世界各国的顶尖棋手正对着这个伟大而逻辑复杂的游戏积极思索。他们究竟在干什么？”哈萨比斯有一种顿悟的感觉，他意识到了这样的游戏完全是劳而无功的。后来，在英国广播公司（BBC）的采访中，他承认当时在想：“我们是不是在浪费大脑，因为这个级别的比赛选手都是顶尖的，为什么不用我们的脑力去做点更有意义的事情呢？比如解决癌症问题，或是找到其他疾病的治愈方法，那样不是更好吗？”

在与他的对手——成年的荷兰世界冠军激战10个小时后，哈萨比斯以微弱的劣势输掉了比赛。赛后他宣布自己将不再参加国际象棋比赛，这让他的父母非常震惊，因为所有人都以为这将是他的未来。但下棋的那些年他也收获颇丰，比如他用击败美国对手亚历克斯·张（Alex Chang）赢得的200英镑奖金，购买了人生中的第一台电脑ZX Spectrum。正是这台电脑促使他产生了要让机器学会思考的想法。

哈萨比斯很快又升级了他的装备——一台由康懋达（Commodore）公司发布的、著名的Amiga计算机<sup>[2]</sup>。闲暇之时他在这台电脑上编程，开发喜欢的游戏。对当时的他而言，开发国际象棋程序还太过复杂。因

此，他开发了一款游戏——奥赛罗（Othello），并取得了成功。奥赛罗是一个黑白棋游戏，当某种颜色的棋子困于相反颜色的棋子之间，其颜色就会反转。这个下里巴人的游戏没办法登上大雅之堂，因此也就没有专业的高手关注和挑战，哈萨比斯只好找他的弟弟来测试程序。结果是程序百战百胜。

该程序采用经典的“如果.....那么.....”选择结构设计，需要编写代码以应对对手的每一个动作，比如对手走了“那一步”，就用“这一步”回击。所有的创造力均来自哈萨比斯本人对于赢得比赛的预判能力。犹如魔法师的徒弟掐诀念咒一般，只要代码编写正确，计算机程序就可以毫无悬念地取得胜利。

看过杰夫·高布伦（Jeff Goldblum）主演的电影《生命的故事》（又名《双螺旋竞赛》）后，哈萨比斯对剑桥大学非常向往：“这真的是在剑桥发生的事吗？就在附近的小酒吧里提出DNA双螺旋学说，这简直太酷了！”所以，在16岁那年，他就向剑桥大学提出申请，想成为一名攻读计算机科学专业的学生，而跟他同龄的其他人此时还在辛苦地学习高中课程。

虽然剑桥大学破格录取了他，但由于年龄太小，学校要求他晚一年入学。为了打发时间，哈萨比斯参加了英国游戏杂志Amiga Power举办的游戏设计大赛并获得了二等奖，也因此得到了去游戏开发公司工作的机会。在那里他开发了一款新游戏“主题公园”，玩家可以打造并经营自己的主题公园。这款游戏获得了巨大的成功，不仅销售量达到了数百万套，还获得了“金摇杆奖”。哈萨比斯也因此赚到了足够多的钱以供自己上大学使用，之后他便动身前往剑桥。

在剑桥学习的时光里，哈萨比斯领略了人工智能领域伟大的科学家艾伦·图灵的风采，知晓了他著名的图灵智能测试；了解了亚瑟·塞缪尔（Arthur Samuel）和他的跳棋程序；提出“人工智能”这个概念的约翰·麦卡锡（John McCarthy）以及弗兰克·罗森布拉特（Frank Rosenblatt）和他的第一个神经网络试验。哈萨比斯渴望站立于巨人的肩膀之上，在人工智能领域有所建树。可是在课堂上教授却反复强调：“围棋极具创造性和直觉性，计算机永远下不好围棋。”这对年轻的哈萨比斯来说就像斗牛比赛中那块猩红的斗篷，激起了他的斗志。所以，当他从剑桥毕业时，他决心通过自己的努力来证明教授的言论是错误的。

他的想法是，与其编写一个会下围棋的程序，不如编写一个通用性



的“元程序”，它可以用于编写出会下围棋的程序。这或许有点不可思议，但重点是“元程序”在实现以后将具有模式学习能力，随着经历的棋局越来越多，该程序会在下棋过程中自我学习，不断地从错误走法中总结经验并加以改进。

哈萨比斯了解到人工智能研究员唐纳德·米基（Donald Michie）在20世纪60年代就实现了类似的想法。米基编写了一个名为“MENACE”的算法，该算法可以零基础学习玩井字棋游戏的最佳策略（MENACE代表导出○和×策略的引擎）。为了验证该算法，米基用304个火柴盒代表○和×所有可能的布局；每个火柴盒里都装满了不同颜色的球，代表可能发生的动作。球被移走或加到盒子里，用以处分失误或肯定胜利。随着积累的游戏经验越来越多，重新分配球的过程最终产生了几乎完美的游戏策略。哈萨比斯打算借鉴这种从错误中改进策略的思路来训练下围棋的算法。

哈萨比斯为算法的实现精心设计了一个巧妙的模型。新生儿的大脑并没有预先设定应对生存挑战的方法，但他们会通过不断学习来强化自我，根据环境的变化做出适当的调整。

哈萨比斯打算研究人类的大脑是怎样学会解决问题的方法的，因为他觉得了解大脑的工作原理有助于实现自己创建一个会下围棋的计算机程序的梦想。因此，他去伦敦大学学院（UCL）攻读了神经科学博士学位。在试验工作闲暇之余，哈萨比斯与神经学家谢恩·莱格（Shane Legg）交流过创办一家实现自己想法的公司。尽管当时人工智能的学术地位很低（甚至在十年前，他们都不曾向教授们表露过想要投身于人工智能事业的想法），但他们仍然觉得自己正在从事一项伟大的事业。2010年9月，这两位科学家与穆斯塔法·苏莱曼（Mustafa Suleyman，哈萨比斯从小一起长大的好友）三人创建了公司，即DeepMind。

在DeepMind成立之初，筹集所需的资金异常困难，因为大多数投资人都认为投资一个研究“人工智能玩游戏”的项目太不靠谱。只有埃隆·马斯克（Elon Musk）、彼得·蒂尔（Peter Thiel）等极少数的投资人看好这家公司的前景并注入了资金。蒂尔从未给硅谷以外的公司投资过，他试图说服哈萨比斯把公司搬到美国西海岸，以便更好地发展。但哈萨比斯是土生土长的伦敦人，他坚信家乡还有很多能够壮大自己事业的人才未被发掘。哈萨比斯还记得他与蒂尔的律师的一次针锋相对的谈话——蒂尔的律师讥讽地问道：“伦敦有关于知识产权保护的法律吗？”他则回斥道：“我想他们以为我们也来自廷巴克图<sup>[3]</sup>！”这场谈论背后的意

义在于，虽然DeepMind创始人不得不把大量原始股份给投资人，但总算是有研究人工智能的资金了。

一开始就创造出一台能够下围棋的机器，这种一蹴而就的想法当然是痴人说梦。所以，他们在开始阶段选择了一个相对简单的目标：20世纪80年代的雅达利（Atari）游戏。在20世纪70年代末到80年代初，雅达利非常流行，很多学生因沉迷这款游戏机而荒废学业。我的朋友就有一台雅达利2600游戏机，他把大量的时间花费在玩Pong、Space Invaders和Asteroids等游戏上。这款游戏机是第一批能够通过软盘加载多款游戏的游戏机，可以直接加载后续开发出来的不同类型的新游戏。再早以前的游戏机只能玩那几款“写死”在硬件上的游戏。

我最喜欢的一款雅达利游戏叫作Breakout（打砖块）：屏幕上方是彩色砖块砌成的砖墙，通过操纵杆的左右移动来控制底部的球拍，小球经球拍反弹后撞向砖墙，被撞到的砖块会消失。撞到砖墙底部的黄色砖块得1分，撞到砖墙顶部的红色砖块得7分。为提升游戏的复杂度使玩家获得更刺激的游戏体验，随着被清理的砖块的数量增多，球拍的尺寸会逐渐缩小，小球的速度也会逐渐加快。

某天下午发生了一件令人非常开心的事情：我们发现了一个破解游戏的巧妙方法。如果能在砖墙的一侧敲出一条通道，一旦小球经该通道到达砖墙顶部区域，它就会沿着通道两侧的墙壁一路反弹，依次清理剩余的砖块。玩家坐等小球自己“工作”完毕从墙上掉落下来后，再用球拍把球反弹出去即可。重复这一动作就可轻松得到高分。这个策略简直太棒了！

哈萨比斯团队的成员在年轻时也花费了不少时间玩电脑游戏。他们的父母要是知道他们为那些游戏投入的时间和精力变得有价值一定会非常高兴。事实证明，打砖块游戏是一个完美的测试用例，可以检验DeepMind团队是否具备开发能够学会玩游戏的程序的能力。对他们而言，只针对一款游戏编写程序有些太简单了，所以哈萨比斯团队又设定了一个更大的目标。

他们想编写一个根据当前屏幕像素状态和得分这两个输入参数之间的关系，寻求最佳得分策略的程序。该程序不会预先设定游戏规则，而是通过随机选择不同的“动作”（比如在打砖块游戏中移动球拍或是在Space Invaders游戏中发射激光炮射击外星人）不断试验，对相应的得分情况进行评估，分析其结果是有效提升还是止步不前。

该程序的实现基于20世纪90年代提出的强化学习（reinforcement learning）思想，目的在于根据分数的反馈或奖励函数来调整执行动作的概率。比如在打砖块游戏中，决定底部球拍左移或右移的初始概率是50：50。如果某次随机移动后，球拍击中小球并且得分，那么程序将根据这些信息重新校准球拍左右移动的概率，以增加球拍击中小球的机会。新的算法将强化学习与神经网络相结合，后者将评估像素的状态以确定哪些特征与加分有直接关系。

刚开始时，程序尝试随机移动，但收效甚微，几乎没有得分。当某次移动后成功得分，程序就会记录这次移动并在后续游戏过程中不断强化该移动的使用频率。一段时间后，随机动作会逐渐消失，代替它们的是一组更为有效的移动。程序在不断试验的过程中，可以真正学会通过特定的移动来提高它在游戏中的得分。

DeepMind团队提交的论文所附带的程序学习玩打砖块游戏的视频很值得一看。刚开始，程序会随机地左右移动球拍，并“观察”会发生什么。当小球被球拍反弹向上并撞击砖块得分时，程序开始重写。它发现让小球的像素与球拍的像素相连接（小球撞击球拍）似乎是一件好事（获得加分）。经过400局游戏的“学习”，程序已经可以精准地移动球拍，并不断地来回击球得分了。

当该程序玩过600局后，它所展现出来的水平令人惊讶！它找到了我们之前破解游戏的那个巧妙办法。这引发了我的感慨：虽然不记得我和小伙伴们玩了多少局才找到这个“窍门”，但从耗费的时间来看，显然我们用的更多。这表明现在计算机程序不仅做到了，而且还做得更快、更好。程序操纵球拍在砖墙的一侧建立通道，使得小球能够抵达砖墙的最上方，在砖墙顶部和屏幕顶部之间的空隙里来回碰撞得分。分数快速增加的同时，程序却无须太多干预。记得当年我们发现这个技巧后兴奋地击掌相庆，可程序轻易就做到了。

时间一晃，到了2014年，也就在DeepMind成立4年后，该项目在已经涉足的49款雅达利游戏的29款中获得了优于人类玩家的表现。该团队在2015年初向《自然》杂志提交的论文中详细介绍了他们的研究成果。众所周知，在《自然》杂志上发表论文是科学家在科研事业上的重要里程碑，可DeepMind团队的论文不仅获得了极高的赞誉，还登上了杂志的封面。《华尔街日报》评论道：“这是人工智能发展史上的重要时刻。”

在这里我必须重申，这是计算机编码的一项伟大进步！以像素状态和不断变化的分数作为原始数据，从随机移动球拍到发现在砖墙两侧建立通道以获取更多得分，程序实现了自我进化。但雅达利游戏的复杂性不可与古老的中国围棋同日而语，所以，哈萨比斯及他的团队决定研发一款可以挑战围棋的更厉害的程序。

此时，考虑到未来的发展，哈萨比斯决定将公司卖给谷歌。他在接受《连线》（Wired）杂志采访时解释说：“本来我们并不想这么做，但在过去3年里，为了筹措资金，我只有10%的时间用于研究。所以，我意识到，我的人生可能没有足够的时间，既能把公司发展成谷歌那样的规模，又可以在人工智能领域有所建树。这样的选择对我来说并不难。”这笔买卖对于谷歌的财力来说可谓轻而易举，同时，也给DeepMind团队研发会下围棋的程序带来了巨大支持和帮助。

[1] 意指不循规蹈矩。——译者注

[2] 非正式译名为阿米加。此计算机为高分辨率、快速的图形响应、多媒体任务，特别是游戏做了专门的设计。——译者注

## 首战告捷<sup>[1]</sup>

此前人们开发的下围棋程序，甚至很难与业余的优秀围棋选手相匹敌。所以，大多数权威人士并不看好DeepMind团队，不认为他们能研发出具有世界冠军水平的专业围棋程序。1997年“深蓝”在国际象棋比赛中战胜了人类棋手之后，大多数人仍然支持天文物理学家皮特·哈特

（Piet Hut）在《纽约时报》上发表的观点：“在围棋领域，计算机想要击败人类可能还需要100年甚至更长的时间。让一个智商正常的人学习下围棋，即使他没有卡斯帕罗夫（Kasparov）的天赋，但几个月后他就能打败目前存在的所有计算机程序。”

DeepMind团队则坚信“一百年太久，只争朝夕”，或许只需要短短的20年就能攻破技术上的瓶颈。他们赋予程序算法的自学习和自适应能力似乎奏效了，但他们仍不能确定新算法到底有多强大。因此，2015年10月，他们决定组织一场非公开的人机对弈来测试程序，对手是当时的欧洲冠军——来自中国的樊麾。

比赛中，新研发的程序AlphaGo以5：0的战绩战胜了樊麾。但欧洲棋手的实力与东亚棋手相差太大：在世界围棋比赛中，欧洲顶级选手只能位列600名左右。因此，虽然这次比赛的成绩令人印象深刻，可这就好比制造出一辆无人驾驶汽车然后在银石赛道上击败了人类选手驾驶的福特嘉年华，并不意味着它能在F1大奖赛中战胜刘易斯·汉密尔顿（Lewis Hamilton）。

当东亚地区的媒体得知樊麾在人机对弈中失利后，它们毫不留情地抨击了这次比赛，并认为AlphaGo的胜利并无多大价值。当消息传出后，樊麾的妻子曾联系伦敦的丈夫劝他不要上网关注此事。显然，他的同胞并不认可他的围棋水平，以及他具备代表人类参加人机对弈的资格。对他而言，这绝对不是愉快的经历。

但樊麾认为，与AlphaGo的对弈使他对围棋有了新的领悟。几个月后，樊麾的世界排名就从633名上升到了300名。事实上，不仅樊麾的实力得到了提升，而且AlphaGo的每一次对弈都会促使它优化代码以应对下一次对弈的挑战。

正因为如此，DeepMind团队才有足够的信心让AlphaGo向拥有18个

世界冠军头衔的韩国棋手李世石发起挑战。

2016年3月9日至15日，万众瞩目的人机大战正式拉开帷幕——李世石九段对阵谷歌AlphaGo五番棋赛在首尔四季酒店举行。比赛通过互联网直播，获胜者将获得100万美元的奖金。为保证比赛不受外界干扰，虽然比赛地点是公开的，但在酒店内的比赛现场却是封闭和保密的。事实上，媒体和现场观众的任何行为都不会让AlphaGo分心，因为机器无论在什么状态下，都会保持“禅宗大师”一般的定力，呈现出一种完美的专注状态。

李世石并没有因为要与打败樊麾的机器棋手对战而感到紧张。在听闻樊麾失利的消息后，他曾表示：“基于AlphaGo在这场比赛中展现出来的水平，我有信心轻松碾压这台机器棋手！”

虽然李世石也知道挑战他的机器棋手可以通过不断的学习和进化变得更为强大，但起初他并不在意。可随着比赛临近，他开始思考人工智能是否最终会强大到可以战胜人类，即使是在围棋这种相当复杂的决策领域。是年2月他曾表示：“听说DeepMind的人工智能异常强大，并且还会通过学习和进化变得越来越强大，但我坚信至少这次对弈我还是能够胜出。”

大多数人认为，尽管人工智能取得了巨大的突破，但想要使拥有人工智能的机器棋手成为围棋世界冠军依然是一个遥不可及的目标。Crazy Stone是唯一一款接近高水平棋手的围棋程序，它的创作者雷米·库仑（Remi Coulom）预测：“还得再有10年时间，才会有电脑打败人类围棋高手。”

大赛在即，DeepMind团队认为还需要继续优化AlphaGo，所以他们邀请了樊麾在大赛前的几周对AlphaGo进行反复的强化训练。尽管遭遇了滑铁卢并被无良媒体攻击，但樊麾依旧乐于支持这项事业。也许在他看来，如果能帮助AlphaGo打败李世石，他就能挽回些许颜面。

在与AlphaGo的对弈训练中，樊麾发现，AlphaGo在某些方面确实非常强大，但此次集训也成功暴露了一个DeepMind团队此前没有意识到的弱点：在某些特定参数配置下，AlphaGo似乎完全无法评估出到底是谁掌控了比赛，常常会产生一种错觉，以为自己赢了，而实际情况却恰恰相反。如果李世石发现并利用这个弱点，AlphaGo不仅会失利，而且还会显得极其愚蠢。



接下来的日子里，DeepMind团队废寝忘食地工作，很快修复了这个缺陷——他们的程序终于可以定版了。在程序调通后，运行该程序的笔记本电脑被运往首尔参赛。3月9日，对弈双方“棋手”（严格意义上讲，只有一名棋手）已经做好充分的准备，迎接人机对弈的五番棋首局。

[1] 原文为“First blood”。——译者注

## 精妙绝伦

多年来，我一直把数学研究与下围棋做比较。所以，我怀着相当焦虑的心情在YouTube上观看了李世石与AlphaGo的比赛。这场“人机大战”吸引了全球超过2.8亿人次在线观看。

李世石执黑先行，作为白方的DeepMind团队由其成员黄士杰（Aja Huang）代替AlphaGo行棋。这样安排是因为，毕竟AlphaGo只是人工智能程序而不是能够自己下棋的机器人。黄士杰紧紧盯着AlphaGo的屏幕，等待它对李世石的落子做出回应，但时间过了一小会儿，AlphaGo没有任何反应。

我聚精会神地盯着屏幕，想知道AlphaGo是不是已经崩溃了！DeepMind团队也开始担心，想知道这是怎么了。对于人类棋手来说，刚开局的几步棋通常都是标准布局套路，是不会想太多的，毕竟还没有进入真正的交锋时刻。在焦急的等待过后，电脑屏幕上出现了一颗白子，AlphaGo终于开始走棋了。直至此时，DeepMind团队才算是松了一口气，心口的大石终于落了地：它可以自行工作了，没我们什么事了！在接下来的时间里，黑白双方你来我往，好不快意。

作为围棋外行的我在观看围棋比赛时会关注一个问题，就是如何判断胜负手，这是围棋比赛的要点之一。这也正是计算机程序学习下围棋非常困难的主要原因之一，因为到目前为止，还没有一种简单易行的方法可以建立起一套稳妥的系统，去评价对弈双方的领先状况。

相较之下，国际象棋更容易进行得分评价，它的比赛规则确定了每一步都会有不同的分值，这样你就可以简单地算出双方的领先状况。国际象棋是破坏性的，在行棋过程中，棋子会被一个个吃掉。正因为如此，棋局会逐步简化。而围棋则不然，它是建设性的，行棋越多，棋盘上的棋子越多，棋局也越来越复杂。现场解说员虽然一直关注着棋局，但不到最后一刻他们也很难断言谁胜谁负，鹿死谁手。

研究团队很快摸清了李世石的开局策略。李世石采用的对战策略是，如果AlphaGo是通过对过去的棋局进行学习来完善自身的，那么他就使用非传统的棋路来打乱AlphaGo的预判，这样就会使他获得一定的优势。但问题是，在这种情况下，李世石也只能按照自己不熟悉的棋路去行棋、布局。对于一般的基于开放式数据库的程序来说，不按套路出

牌的策略非常管用：不仅可以使机器手足无措，还可能误导机器在棋局的重要关口或是长远战略决策上犯下致命错误。但遇到AlphaGo，这个如意算盘可就打空了，因为AlphaGo可以实时动态评估棋局形势，并根据以前的经验制定出最佳策略。AlphaGo的首席程序员大卫·西尔弗（David Silver）在赛前表示：“AlphaGo可以自动把数以百万计的棋局数据通过神经网络进行深度学习并加以改进，以得到最佳的对弈策略。相比之下，李世石则由于无法掌控自己的棋局而会变得越发被动。”

看到这里，我不禁同情起李世石来。当他意识到自己一直处于被动时，你能真切地感觉到他的信心正在逐渐丧失。他一直盯着代替AlphaGo行棋的黄士杰，但却从对方的脸上看不到任何线索。当下到第186手时，李世石感到大势已去，便投子认输了。

五番棋首局后，AlphaGo以1：0暂时领先。李世石在当天的新闻发布会上表示：“我很惊讶，我从未料想到自己会输给AlphaGo。”

第二局才是好戏一场，不仅仅让参赛的李世石瞠目结舌，还让观看比赛的棋手们感到匪夷所思。第一局比赛中AlphaGo所走的每一步棋还是符合人类逻辑思维的，现场的专家也能够讲解和分析棋局。但到了第二局，“事出反常必有妖”，当李世石下完第36手来到酒店屋顶抽烟休息时，AlphaGo下出第37手：黄士杰在距离棋盘边缘5步的位置落下一颗黑子。这一招使得包括李世石在内的所有人都倍感震惊。

根据传统的行棋套路，在初始阶段，应先在最外面的4条线上落子，这样就可以在棋盘边缘向内的第3条线上形成短效的区域控制力量，而在第4条线上落子则可以帮助棋手在棋局的后期向中心扩张自己的势力范围。所以，棋手们一般会选择先在第3条线或第4条线上落子，以达到一种较好的攻守平衡。在第5条线上落子一般被认为是不太恰当的选择，因为这会给对手可乘之机：建立一个既可在短效、局部区域内抢得先手，又可在长远、全局范围内影响胜负的策略。

此步棋打破了长久以来的传统观念，有评论员抨击这种下法明显是胡闹，但也有一些评论员持谨慎态度。大家都很期待看到李世石的应对策略。当李世石重返赛场看到此棋时，他在盒中取子的身体在往后缩……他举棋不定，坐在那里思考了超过12分钟。围棋和国际象棋同样都限定了落子的时间，用12分钟来思考一步棋的代价很大。李世石花费了如此之久的时间才落子，也从侧面说明了AlphaGo的这步棋是多么的出人意料。他弄不明白AlphaGo到底想要干什么：为什么程序主动放弃

了它正在“争夺”的区域？

难道是AlphaGo出现了错误？还是它发现了棋局中被李世石忽视的某些东西？当担任裁判的樊麾看到这步棋时，他最初的反应与其他人一样，除了惊讶就是惊讶。但是，他很快意识到了，这确实不是人类的行棋方法，他从未见过有人会这样行棋。他不禁感叹：“这步棋真是神来之笔，精彩绝伦！”接下来的比赛证明，这一步既精彩又致命：这一着非但不是臭棋，反而是立意深远的妙手。大约又过了50子，黑白双方开始在棋盘的左下角争夺地盘，棋路向AlphaGo第37手时落下的那颗黑子靠近。正是由于AlphaGo的这一预先布局，让黑方在这一片区域赢得了优势。正是这制胜的一子，让AlphaGo以2：0继续领先于人类。

在当日赛后的新闻发布会上，李世石的情绪明显不同于前日：“昨天的我很惊讶，但今天的我无言以对，心中满是震惊。坦白地说，第三场比赛对我来说将不是那么容易了。”大家都很清楚，这场人机对弈共有五局，从第三局开始李世石必须全胜才有可能逆转战局、反败为胜。

## 人类的反击

第三局比赛在3月12日星期六举行，李世石有一天的时间可以用来休息、恢复状态。第一局比赛进行了三个多小时，第二局进行了四个多小时。在这方面，人与机器毕竟不同，人需要休息，而机器是不需要的。连续的两场败局无疑给李世石带来了很大的精神压力和挫败感。

然而，李世石并没有躺下来睡大觉，而是与一群职业棋手分析和探讨此前两场比赛中失利的原因，这场讨论一直持续到次日清晨6点。AlphaGo有弱点可以利用吗？李世石认为不只机器可以学习和进化，人也可以从失败中学到一些东西。

第三局比赛开局时，李世石非常强势，并取得了不错的战果。在他凌厉的棋势下，AlphaGo只能被驱赶，被压迫在一小片区域内。这种一边倒的局面使得评论员都开始兴奋了起来，有人直接表示李世石已经发现了AlphaGo的弱点。但在后来，局势出人意料地急转直下，一位评论员回忆道：“形势急转直下，眼睁睁地看着所发生的一切，让我深感不适……”当李世石把AlphaGo逼到绝境，AlphaGo的巨大潜力似乎被激发出来了。随着比赛的进行，AlphaGo开始采用被评论员称为怠惰走法（lazy moves）的策略。通过分析，AlphaGo确信自己最终可以获胜，正因为如此，它选择了这种安全的策略。它并不关心能赢多少目，重要的是它最终一定会赢。这种耍赖一样的怠惰走法可能对李世石来说有些冒犯，但AlphaGo这种做法并没有任何挟私报复的性质。它这样做只是单纯地为了赢棋。李世石不想就此认输，他在棋盘上坚持应对着，觉得也许在这些怠惰走法中还有可乘之机。

但在第176手后，李世石还是投子认输了，AlphaGo以3：0战胜了人类。不过在后台，DeepMind团队的成员却产生了异样的情绪：虽说他们已经赢得了比赛，但他们看到的AlphaGo对李世石造成的巨大影响使他们很难高兴起来。百万奖金很显然已是他们的囊中之物，在比赛之前DeepMind团队就已经达成共识，如果能赢得这笔奖金，就将它捐赠给致力于推广围棋和科学研究的项目以及联合国儿童基金会等慈善机构，但人类的思维和逻辑让他们对李世石的痛苦产生了同情。

AlphaGo获胜后没有任何情绪反应，不会产生任何小小的电流波动，更不会显示出一段代码或是大叫一声“爽！”。机器的这种“镇定”给

人带来希望的同时，也让我们为将来感到忧心。带来希望，是因为正是人类的这种情绪反应激励着我们去探索未知、开创未来，毕竟还是人类给AlphaGo编写了制胜的代码。感到忧心，是因为机器太过“冷漠”，它根本就不关心事情发展的最终结局是不是程序编写者所期望的。

李世石被三连败的事实打击得有些恍惚，他在新闻发布会上道歉说：

不知该如何开口，也不知道今天要说些什么，我必须为辜负了大家的期望表达歉意。我应该在比赛中表现得更好，我尽力了，但我还是体会到了深深的无力感。

同时，他也呼吁大家继续关注后续的两场比赛，他现在的目标是至少为人类扳回一局。

连输三局，整个比赛的胜负已定。李世石在第四局时似乎卸下了沉重的思想包袱，放松了许多，这样反而使他能够更专注于比赛本身。在第四局中，与第三局慎之又慎的行棋风格形成鲜明对比的是，李世石采用了一种更为激进、极端的“先捞后洗”（amashi）的策略。一位评论员将这种冒险的行为比作：一位金融家，不想通过日积月累的小收益来积累财富，而是激进地将整个银行作为赌注押了出去。

李世石和他的团队在星期六晚上通宵达旦，试图通过逆向工程的方法掌握AlphaGo的下棋思路。他们发现AlphaGo似乎遵循着逐步增加获胜概率并去调整落子的原则，而不是押注于复杂的单一走法。在第三局中，李世石就亲身体验了AlphaGo的怠惰走法。他们针对这个特点制定了应对策略，通过一着险棋来扰乱AlphaGo的策略：“胜负手”（all-or-nothing）策略可能会让AlphaGo更难轻易得分。

AlphaGo似乎对这种策略毫不在意。比赛进行到第70手时，评论员开始觉察到AlphaGo再次占据了上风——当AlphaGo的棋路开始变得保守，频频使用怠惰走法时，就标志着AlphaGo已经确认自己领先了。李世石必须要做点什么的，否则很难扭转乾坤。

如果说第二局对弈的第37手是AlphaGo的神来一笔，那么第四局的第78手就是李世石的逆袭大招。当他盯着棋盘审时度势了30分钟后，突然在看似不寻常的一处落下白子，刚好在AlphaGo的两个黑子之间。当时在YouTube上发表评论的迈克尔·雷德蒙德（Michael Redmond）说



道：“这一步让我感到意外，我相信大多数人，包括AlphaGo，都会觉得措手不及。”

AlphaGo确实有些无措，下了一步莫名其妙的棋作为回应。又经过了几手棋，AlphaGo就发现自己败局已定。紧盯着电脑屏幕的DeepMind团队发觉他们的程序在第78手时就开始失控了。之后，AlphaGo已经彻底崩溃，所以下了一连串的昏招。这显然是围棋程序的另一特点：当它意识到自己失败后，会做出一些令人费解的疯狂行为。

西尔弗看到AlphaGo的表现后说道：“我觉得大家看到这一幕会感到很搞笑！”果不其然，韩国的评论员们开始嘲笑AlphaGo的混乱状态。很显然，AlphaGo的行为没有通过图灵测试，因为任何一个具有战略眼光的人都不会做出那样的决策。当下到第180手时，AlphaGo在电脑屏幕上表示投子认输。新闻演播室里爆发出雷鸣的掌声。

人类终于扳回一局，3：1！当晚的新闻发布会上，李世石的笑容说明了一切：“这场胜利是如此的珍贵，世界上的任何东西都无法替代。”伴随着媒体的欢呼，他高声说道：“我能取得这局对弈的胜利是因为你们所有人给我的欢呼和鼓励。”

当时，在中国评论这次比赛的古力称赞道：“李世石的第78手实乃‘上帝之手’。”那一步棋打破了传统棋路，是为整局比赛带来深远影响的关键所在。这是真正的人类创造力的体现，也是博登变革性创造力的一个极好的典范。不破不立，当你突破原有系统的束缚时，你的创意就会层出不穷。

哈萨比斯和西尔弗在新闻发布会上表示无法解释AlphaGo失利的原因，他们需要回去好好分析——为什么它在应对李世石第78手时会走出如此糟糕的一系列昏招。事实上，AlphaGo与人类对弈的历史经验让它完全摒弃了某些思考。根据它的评估，那种下法只有万分之一的可能性会出现。它似乎没有认真考虑去学习如何应对那种下法，所以就不自觉地优先考虑了应对其他下法的可能性。

如果李世石有足够长的时间深入了解对手，那么他在比赛中战胜AlphaGo的可能性或许会增大。他能在第五局比赛中保持获胜的势头吗？同样是输掉比赛，3：2还是4：1的结果大不相同。最后一局意义非凡，如果李世石能拿下第五局的胜利，那么他将在人们的心中播下对“AlphaGo绝对优势”怀疑的种子。

经此一役，AlphaGo也收获颇丰。现在就算李世石在第10 000手下出违反常规的怪招，它也不会再想着侥幸过关了。这就是算法的强大之处：从错误中吸取教训，进而反败为胜。

但这并不意味着AlphaGo不会再犯新的错误。随着第五局棋的进行，AlphaGo似乎放弃了一套针对特定布局的标准下法。正如哈萨比斯在后台所讲到的：AlphaGo在开局之初犯了一个错误（因为它并不知道围棋中有一个著名的“手筋”<sup>[1]</sup>），但现在它正在努力地改正，亡羊补牢。

开局时李世石处于主动，之后AlphaGo渐渐地挽回了颓势。比赛非常激烈，直到比赛结束之前，DeepMind团队也还不清楚AlphaGo是否获胜了。历时5个小时的人机大战，在第281手棋时分出了胜负，李世石中盘认输。这一次后台一片欢呼，DeepMind团队击掌相庆，哈萨比斯激动地握紧拳头挥向空中以庆祝胜利。李世石在第四局的获胜重新激起了他们的斗志，对他们来说赢得最后一局尤为重要。

回顾此次人机对弈，很多人都意识到了它非凡的意义。一些人立即发表了评论：“这是人工智能的重要转折点！”从表面上看，AlphaGo所能做的仅仅是下围棋，但实际上，它的学习和适应能力才是最值得人类关注的一种全新的东西。哈萨比斯在赢得首局胜利后就曾发表Twitter消息：“AlphaGo赢了！我们成功地把它降落在了月球表面。”这是一个很好的比喻，登月并没有产生关于宇宙的非凡的新突破，但却意味着我们为实现这一壮举而开发的技术产生了非凡的新突破。赛后，韩国围棋协会授予AlphaGo荣誉九段称号，这是围棋选手所能获得的最高荣誉。

[1] 围棋术语，指“灵感之下的妙手”。——译者注

## 山头还是主峰

第二局中AlphaGo的第37手棋很有创意，下法之新颖让人惊喜。随着棋局发展，它证明了自己的价值。这是一种具有探索性的创造力，挑战了传统行棋思维的极限。

围棋最重要的一点，是可以通过客观的方法检验新的行棋思路是否具有价值。任何人都可以想出富有创意的新走法，其技巧性和实用性取决于所能带来的某种价值。我们该如何去衡量这种价值？在评估价值的过程中是否会带有主观臆断，或受到时机的局限性的制约？一些在诞生之初就受到指摘的事物，往往经历几代人才会逐渐被接受并被视为具有革命性的创新。并不被19世纪的人们所认可或知晓的贝多芬的交响乐，现在被誉为艺术的巅峰。凡·高的一生中，几乎没有售出过画作，它们只能用来交换食物或绘画材料，但现在他的大作却可以卖出数百万美元的高价。可是在围棋方面，却有一个最为直接和实际的价值衡量标准：能赢棋吗？AlphaGo凭借精妙绝伦的第37手棋最终获得了第二局的胜利，这也是其价值最好的、最客观的体现。

AlphaGo将这种古老棋类游戏的新下法教给了全世界。分析其在比赛中所采用的新战术，我们已经能够理解它在第5条线早早落子对胜负的影响，越来越多的棋手将会把这种策略运用起来。同时，AlphaGo也在不断地学习和创新，探索更多、更好的行棋策略。据DeepMind团队透露，2017年初，AlphaGo的最新版本在网上以假名Master或Magister和一批围棋高手对战，人类棋手并不知道他们在和机器下棋。几个星期的时间里，它共参与了60场比赛，并获得了全胜。

真正的洞察力源于对棋局的综合把控。这些棋类游戏现在已经成为挖掘新思想的宝库。在几场比赛中，AlphaGo使用了一些新手都不会用的低级招式。比如，传统下法中棋手不会在三行三列交叉点上落子，但AlphaGo却向我们展示了如何利用好这一招并为整个棋局带来新的机遇。

哈萨比斯解释了围棋算法是如何陷入数学家们所说的“局部极大值”的困境当中的。如图3-1所示，如果你登上了A峰的最高点，四周弥漫着的大雾阻碍了你的视线，在此处高无可高的情形之下，你或许会觉得这里就是所在区域范围内的最高点。但当迷雾逐渐被风驱散，举目远

眺，你发现山谷对面竟然还有一座更高的山峰——B峰。想要抵达B峰的顶端，你就必须先下山，穿过山谷，然后才能攀上主峰之巅。

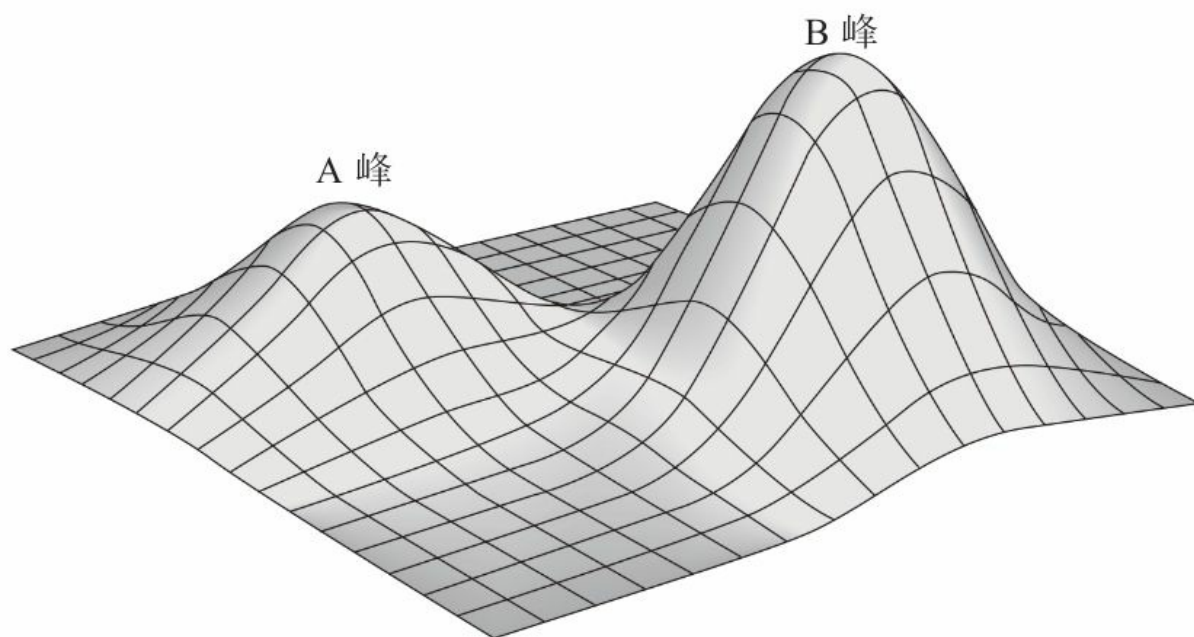


图 3-1

在传统棋法的影响之下，围棋已发展出固有的一套成规，应用好这些规则的确能让棋手登上A峰。AlphaGo的出现拨开了迷雾，打破了这些规则对思维的束缚，使我们看到了更高的B峰。我们甚至可以估算出两者之间的差距。围棋比赛的统计资料显示，使用传统下法的棋手通常会输给使用新下法的棋手两子。

对围棋下法的革新一直持续不断、屡见不鲜。最近一次是围棋界的传奇人物吴清源大师于20世纪30年代开创的新棋法，他的布局之法颠覆了传统围棋布局的常用套路。因此，围棋爱好者们认为AlphaGo可能会引发一场更大的围棋“革命”。

中国的围棋世界冠军柯洁也认为，我们处在一个新的时代，虽然人类已经发明围棋数千年了，但人工智能技术的出现让我们感觉到人类对围棋的理解仍然还很肤浅。在不久的未来，人类与计算机的融合将会开创一个崭新的时代。

柯洁的师兄，另一位中国围棋世界冠军古力补充道：“人类将利用人工智能技术很快揭开围棋更深层次的神秘面纱。”哈萨比斯把人工智

能算法比作哈勃望远镜，认为它是一种可以用来探索比以往更深、更远、更广领域的工具——它会提升而不是取代人类的创造力。

此时此刻，我似乎发现了一个令人相当沮丧的事实。当你知道永远无法战胜机器时，成为围棋世界冠军的理想就会变得毫无意义。专业棋手都试图勇敢地面对这一局面，讨论并分析人工智能所展现的创造力能否对自己棋力的精进大有裨益。但当我们得知再怎么努力也只能成为屈居于机器之后的第二梯队棋手时，确实会意志消沉。虽然机器的程序还是人编写的，但这也不会让人有挽回颜面的感觉。

自那以后，AlphaGo就退出了围棋竞技领域，DeepMind围棋团队也宣布解散，因为哈萨比斯已经用事实证明了剑桥那个教授所持观点的错误性。DeepMind团队现在把目光投向了其他领域：医疗保健、气候变化、能源效率、语音的生成和识别、计算机视觉，等等。这些领域都需要他们认真、深入地思考和探索。

鉴于围棋一直作为我反对利用计算机研究数学的挡箭牌，我关心的是DeepMind团队关注的下一个目标会不会是我所从事的数学研究领域？要正确判断这种新的人工智能技术具备的潜力，我们需要更加仔细观察并深入研究其运作的方式和机理。试想一下，DeepMind团队将要开发出来的程序居然有可能让我这个数学家丢掉饭碗，而创造这些程序的工具正是数学家们历经几个世纪的不懈努力才发现和创造出来的。这个数学缔造出来的“科学怪人”（Frankenstein's monster）会向它的创造者宣战吗？



## 第4章 算法，网络时代的生活秘诀

阿达·洛夫莱斯

分析引擎编制代数的模式，就如同用提花织布机编织出鲜花和绿叶一般。

当下，我们的生活完全依赖于算法。上网搜索内容，使用GPS导航，观看由奈飞公司（Netflix）推荐的电影，或是在线预约，我们都要依赖算法。算法正在引导我们进入数字时代，但很少有人意识到在计算机诞生之前，算法作为数学的核心已存在了数千年。

自古希腊以来，算法的身影一直伴随着数学的诞生与发展。在欧几里得（Euclid）的巨著《几何原本》<sup>[1]</sup>中，他除了证明素数有无穷多个外，还发现了一个方法，按照这个方法就能解决最大公约数等问题。

这也许有助于我们更清楚地认识 and 解决问题。想象一下，如果你的厨房长36英尺<sup>[2]</sup>，宽15英尺，那么能够覆盖整个地面而无须切割的方形瓷砖是多大尺寸呢？你该怎么计算呢？2000年前解决这类问题的算法是这样的：

假设你有两个数字， $M$ 和 $N$ ，且 $N$ 小于 $M$ 。首先用 $M$ 除以 $N$ ，得到的余数记为 $N_1$ 。如果 $N_1$ 为零，那么 $N$ 就是能够将这两个数整除的最大的数，即这两个数的最大公约数。如果 $N_1$ 不为零，则用 $N$ 除以 $N_1$ ，得到的余数记为 $N_2$ 。如果 $N_2$ 为零，则 $N_1$ 是能将 $M$ 和 $N$ 整除的最大的数。如果 $N_2$ 不为零，则继续上述步骤，用 $N_1$ 除以 $N_2$ 并得到余数 $N_3$ 。依此类推，得到的余数是一个整数，并且随着计算的进行会越来越小，直到变为零。那么，算法最终一定会找到一个能够同时将 $M$ 和 $N$ 整除的最大的数，这个数被称为最大公约数。

现在让我们回到厨房地板的问题。我们知道，厨房是长方形的，而我们要寻找的是正方形的瓷砖。假定我们讨论的是一种理想状态：瓷砖的尺寸不会受到生产厂家某些规格标准的限制。现在，我们可以开始了。首先，我们找到适合原始形状的最大方形瓷砖；然后，我们寻找到适合剩余部分的最大正方形瓷砖，依此类推……剩余的地面空间逐渐缩小，直至成为一个正方形，这时刚好就可以用一整块瓷砖严丝合缝地填



充进去。整个过程不需要切割任何一块瓷砖，如图4-1所示。



图 4-1

我们将上述问题的解决思路（算法）用数学的方式加以描述：假设 $M=36$ 且 $N=5$ ，则用 $M$ 除以 $N$ 得到余数 $N_1=6$ ，用 $N$ 除以 $N_1$ 得到余数 $N_2=3$ ，而 $N_1$ 除以 $N_2$ 根本就没有余数，所以就可以得出3是36和15的最大公约数。

你可以看到整个计算过程隐含有许多类似于“如果.....那么.....”的条件判断句式，这是算法的典型特征，也是计算机程序中算法的妙趣所在。欧几里得的古老方法触及了任何算法都应该具备的四个关键特征的核心：

- （1）它应该由一组精确的陈述和明确的指令组成。
- （2）无论输入的参数如何，这个过程都应该完成（不应该进入无限循环）。
- （3）它应该给输入算法的任何参数以答案。
- （4）在理想情况下，它的运行速度应该很快。

在欧几里得的算法中，任何阶段都不存在歧义。因为余数在每一步运算后都会变小，有限的步数之后它必为零，这时算法就会停止并给出结果。算法的执行时间与问题的规模成正比，数字越大，耗时越长。

如果最古老的算法可以追溯到2000多年以前，那为什么“算法”这一名词的提出要归功于一位9世纪的波斯数学家呢？穆罕默德·阿尔·花拉子密（Muhammad Al-Khwarizmi）是巴格达智慧馆<sup>[3]</sup>（great House of Wisdom）的首批负责人之一，他负责将古希腊数学原著翻译成阿拉伯文。“算法”是拉丁文对他名字的翻译。尽管欧几里得的算法在《几何原本》中早已阐明，但欧几里得所使用的语言非常笨拙，而且古希腊人的思维非常几何化（数字只是线条的长度，就连证明的过程都是由图片组成的——有点像我们用瓷砖铺厨房地板的例子），所以他的算法并没有被后世所广泛采用。这是因为图片并不是一种严谨的数学方法，你需要的是代数的语言，即一个字母可以作为变量代表任何数字，而这正是花拉子密的发明。

你需要一种语言来清楚地表达算法的工作原理，并允许你在不指定数值的情况下讨论数学问题。我们已经看到这种语言能解释欧几里得算法的工作原理，给予数字一个形式化的符号名称——N和M，这些符号可以代表任何数字（我们称之为变量）。这种新的描述语言是一种高度概括的语言，它对数学的发展影响巨大，意味着数学家不需要挨个讨论遇到的所有问题，而可以运用形式化的描述方法来掌握数学运算背后的模式。一个好的算法应该满足上述的第三个特征，就好比代码和程序，它们可以不需要确定具体的参数就能够编译运行。

算法已成为我们这个时代通行的“货币”，因为它们是现代计算机系统的完美素材。算法利用我们解决问题的模式，反过来引导我们去找到解决问题的方案。计算机不需要思考，它只要不停地遵循算法、执行指令即可，就像变魔术一样，答案自己就会蹦出来。

**[1]** 欧几里得为了满足教学的需要编成了一部《几何原本》。这部书共15卷，是一部集前人思想和欧几里得个人创造性于一体的不朽之作。在此书中，欧几里得把人们公认的一些事实列为定义和公理，以形式逻辑的方法，用这些定义和公理来研究各种几何图形的性质，从而建立了一套从定义、公理出发，论证命题得到定理的几何学论证方法，形成了一个严密的逻辑体系——几何学。而这部书，也就成了欧式几何的奠基之作。——译者注

**[2]** 1英尺=0.3048米。

**[3]** 中世纪阿拉伯阿拔斯王朝在巴格达建立的全国性综合学术机构，亦称益智宫、哲理大学，由翻译局、科学院和图书馆等机构组成。后世学者誉智慧馆为中世纪阿拉伯“科学的源泉，智慧的宝库，学者的圣殿”。

——译者注

## 荒岛算法

当代最非凡的算法莫过于每天帮助数百万人浏览互联网的搜索算法。如果我被扔在荒岛上，只允许随身携带一个算法，我可能会选择谷歌的搜索引擎。（并不是说它会有多大的用处，因为我不太可能连上互联网。）

20世纪90年代早期，一个专门收录所有互联网网址的目录直到1994年才收录了大约3000个网站。当时的互联网很小，小到可以让你轻而易举地浏览一下就能找到想找的东西。从那时起，互联网就在不断发展。当我开始写这篇文章时，互联网上已经有1 267 084 131个网站了，还没写几句话的工夫，这个数字就上升到了1 267 085 440（通过访问<http://www.internetlivestats.com/> 即可查看现存网站数量）。

谷歌的搜索引擎是如何从数以几十亿计的网站中准确地找出你想要的信息呢？86岁来自维冈的老奶奶玛丽·阿什伍德（Mary Ashwood）在浏览器的搜索栏里非常有礼貌地使用“请”和“谢谢”来提出她的搜索请求，或许她脑海中想象的画面是，电脑另一端一群勤劳的实习生正在埋头苦干，处理浩如烟海的请求信息。她孙子本（Ben）打开笔记本电脑，看到“麻烦你帮我翻译一下这个罗马数字——MCMXCVIII，谢谢你！”的搜索问句后，忍不住在Twitter上向全世界讲述了奶奶的误解。当谷歌工作人员回复他的Twitter消息时，他惊呆了：

敬爱的本的奶奶，

愿您一切安好。

在数十亿次的搜索中，是您让我们会心一笑。

嗯，您所需要翻译的罗马数字是1998。

感谢您！

在这一次的机缘巧合下，本的奶奶将谷歌的工作人员推向台前。谷歌的搜索引擎每15秒就要处理数百万次请求，这样的数量任何公司都无法做到人工回复。那么，如果谷歌不是拥有互联网神奇魔法的精灵，它是如何成功地找到你想要的答案呢？

这一切归功于1996年拉里·佩奇（Larry Page）和谢尔盖·布林（Sergey Brin）在斯坦福大学的宿舍里发明的强大而精妙的算法。他们最初想把新算法命名为“网络爬虫”（Backrub），但最终还是决定叫“谷歌”（Google），其灵感来自1后面的100个零。他们的目标是找到一种对互联网上所有的页面进行排序的方法，以帮助大家在这个不断增长的海量数据库中进行检索，所以起这个代表巨大数字的名字似乎特别贴切，而且也很酷炫。

这并不意味着其他的算法不能做这件事，但是那些算法在概念上非常简单。如果你想搜索更多关于“有礼貌的奶奶和谷歌”的信息，现有的算法会将所有包含这些关键词的页面识别出来，并按顺序排列，搜索词出现频率最高的网站会被放在最顶部。

这种方式虽然有效，却容易被黑客攻击：任何一个花店老板只要在网页的元数据中数千遍地插入关键词“母亲节鲜花”，那么每个想买花的子女电脑上的搜索结果的最顶端就会出现这个花店的链接。你肯定不希望自己的搜索被精明的人设计或者操纵，那么，如何才能对一个网站的重要性给予公正的评价呢？如何判断哪些网站该被过滤掉呢？

佩奇和布林想出一个聪明的方法：如果一个网站有很多链接指向它，就暗示着其他网站认为这个网站值得访问。其原理是通过其他网站的评估去衡量某个网站的重要性，或者说该网站的访问价值。但是，这种方式也有可能被黑客攻击，比如只需伪造出有1000个网站的链接指向这个花店就行了，这样也会使其被纳入搜索名录。

为了防止这种情况出现，他们决定给那些获得广泛好评，深受信赖的网站赋予更高的权重。

可这仍然会让他们面临一个挑战：如何客观评价一个网站的重要性？

以一个小型网络为例，如图4-2所示。首先，给每个网站设定相同的权重。然后，让我们把网站想象成一个桶，给每个桶里放8个球，表示网站的初始权重相同。现在，每个网站必须将球交给它链接的其他网站，如果链接多个网站，那么就将球均分给那些网站。如图4-3所示，由于网站A链接了网站B和网站C，它将为每个网站提供4个球；而网站B只链接了网站C，它就需要将拥有的8个球全部放入网站C的桶中。第1轮分配后，网站C得到的小球数最多。

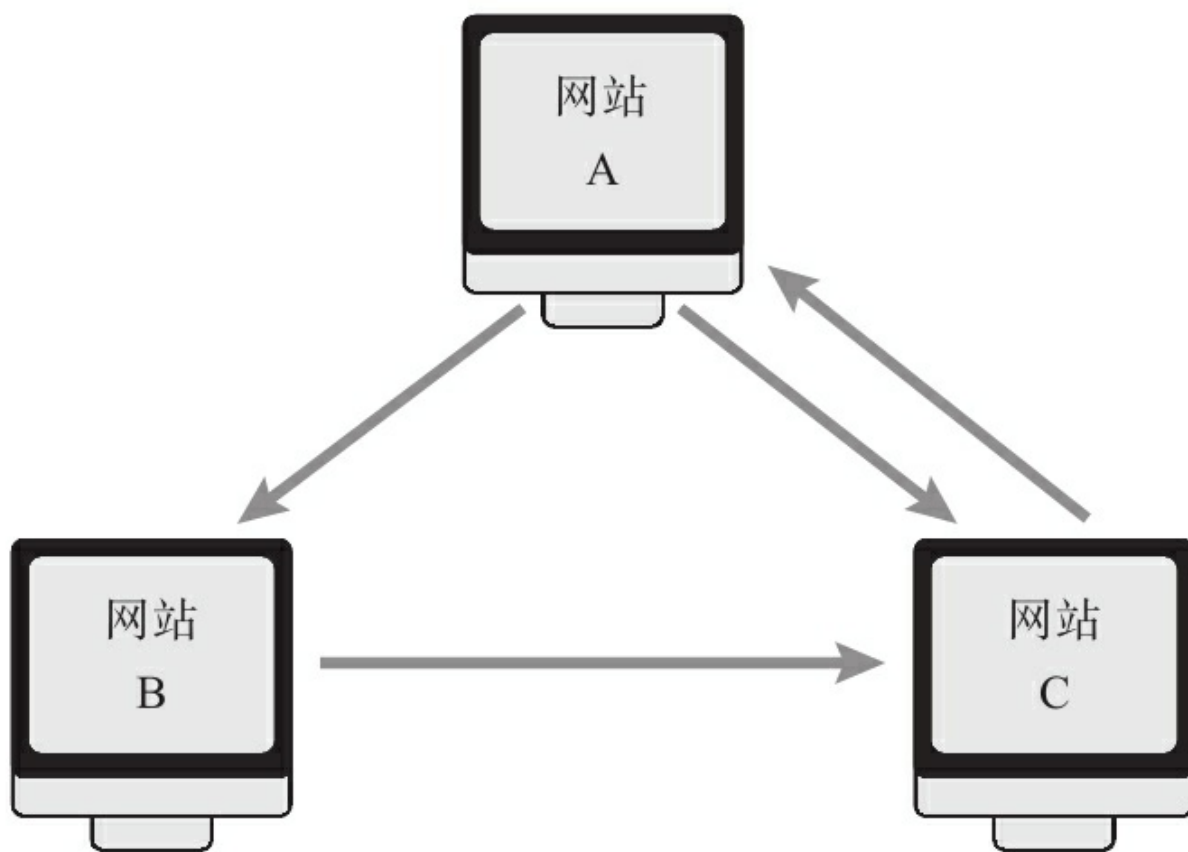


图 4-2

但是我们需要继续重复这个分配过程，因为现在位于最高排名的网站C链接了网站A，所以又会产生新的分配结果。9轮重复分配过程中各网站小球数量的变化情况如图4-4所示。



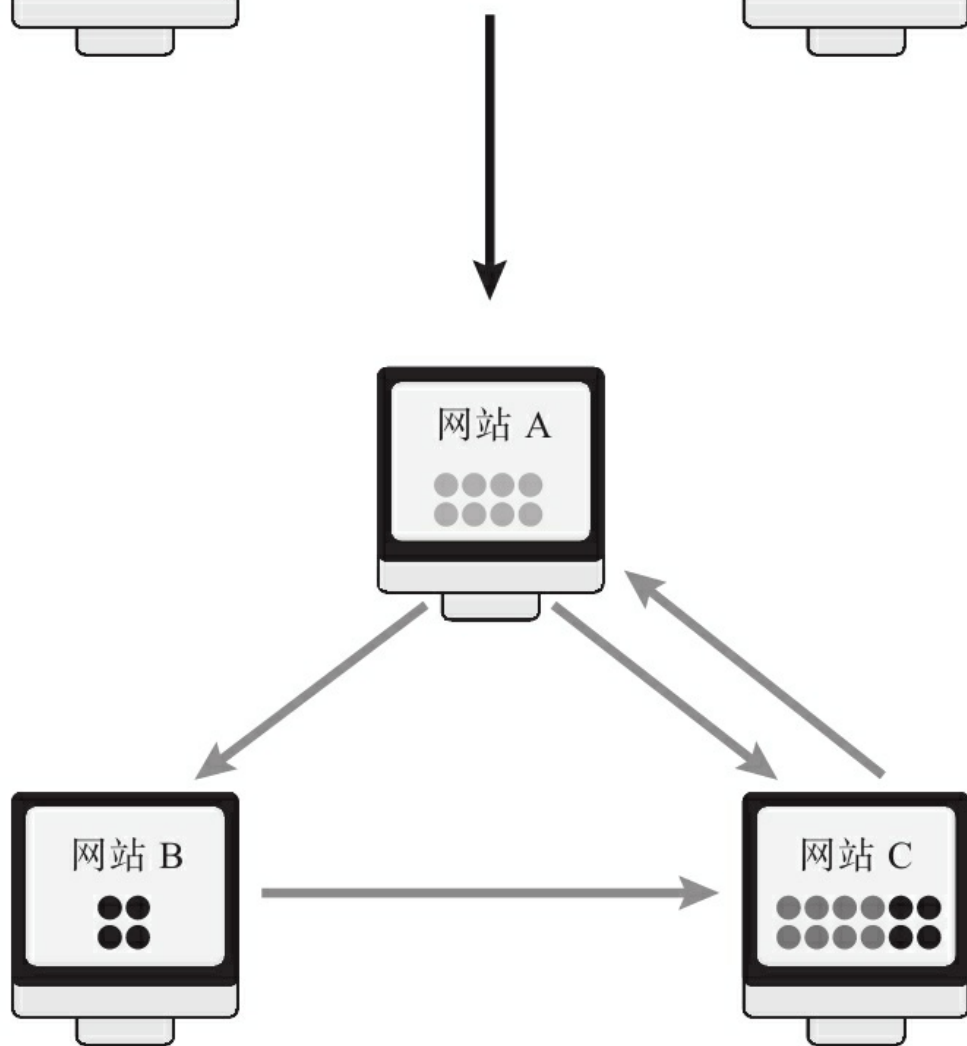
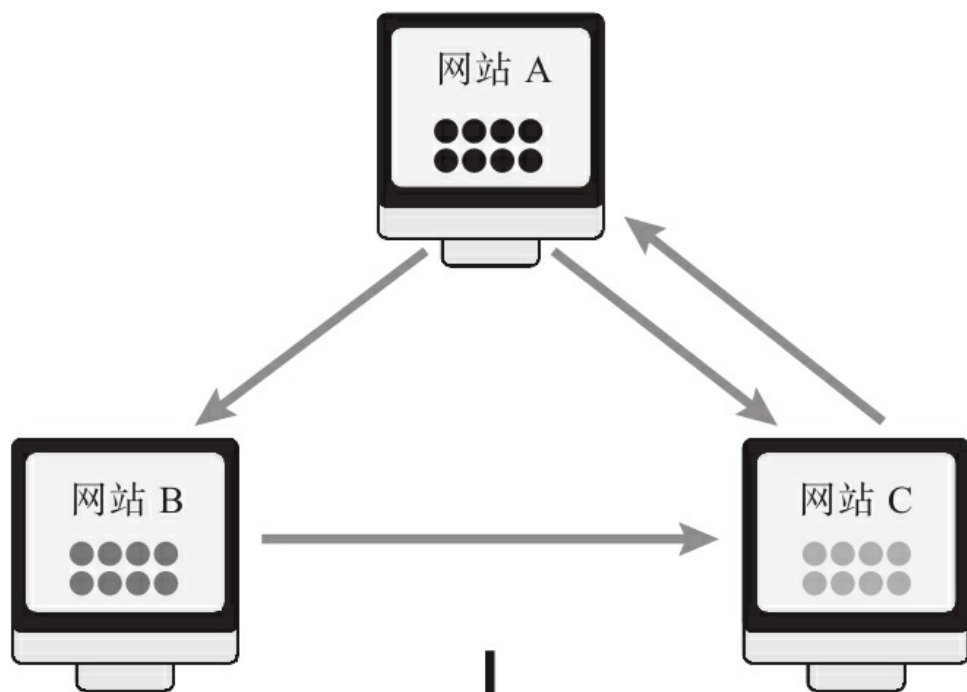


图 4-3

	第 1 轮	第 2 轮	第 3 轮	第 4 轮	第 5 轮	第 6 轮	第 7 轮	第 8 轮	第 9 轮
A	8	8	12	8	10	10	9	10	9.5
B	8	4	4	6	4	5	5	4.5	5
C	8	12	8	10	10	9	10	9.5	9.5

图 4-4

到这一步，它还算不上是一个特别好的算法，因为不稳定，并且效率相当低，没有达到理想算法的两个关键标准。佩奇和布林的洞见之伟大在于，他们意识到，需要找到一种方法，通过观察网络的连通性来分配球。结果，他们在线性代数中找到了一个诀窍，可以一步算出正确的分布情况。

这种算法从构建一个矩阵开始，该矩阵描述球在网站间的重新分配方式。矩阵的第1列表示球从网站A到其他网站的分配比例：0.5转到网站B，0.5转到网站C。由此，可以得到球的重分配矩阵：

$$\begin{pmatrix} 0 & 0 & 1 \\ 0.5 & 0 & 0 \\ 0.5 & 1 & 0 \end{pmatrix}$$

难点是寻找这个矩阵特征值为1的特征向量，这是一个与该矩阵相乘不会发生改变的列向量。找到特征向量的方法我们在大学本科时就学过了，因此在这个网络中我们发现，通过重分配矩阵找到的列向量非常稳定：

$$\begin{pmatrix} 0 & 0 & 1 \\ 0.5 & 0 & 0 \\ 0.5 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$$

注：矩阵的乘法运算规则是：

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} ax + by + cz \\ dx + ey + fz \\ gx + hy + iz \end{pmatrix}$$

这就表明，如果我们按照2：1：2的比例给各网站分配球，会看到这个权重比例是稳定的。用之前9轮分配的例子中得到的数据也可以验证这一结论，各网站拥有的球的比例总是约等于2：1：2。

矩阵的特征向量是在数学和其他科学领域中非常有效的一种工具，是量子物理中用来计算粒子能级的秘密武器，可以用于研究旋转流体的稳定性（比如旋转的恒星或者病毒的繁殖率），甚至可以用于研究素数在所有数字中是怎样分布的问题。

通过计算网络连通性的特征向量，我们发现网站A和网站C的排名应该是相同的。虽然网站A只连接到一个网站（网站C），但由于网站C的权值较高，它会赋予网站A较高的权值。

这是算法的核心基础，但需要加入一些额外的细节处理才能使其充分发挥作用。例如，该算法可能需要考虑一些异常情况：如果存在未链接其他网站的孤立网站，它的球会无法重新分配。

尽管基础引擎是公开的，但算法内部的一些重要参数还是保密的，并且随着时间的推移不断发展变化、更新换代，这些在一定程度上使得算法难以被破解。谷歌算法最吸引人的地方在于它本身的健壮性和防止欺骗的策略——一个网站很难在自己的网站上做手脚来提高排名，它必须依靠其他网站来提升自己的排名。

如果你关注一下谷歌搜索，就会发现排名很靠前的网站主要都是新闻媒体网站和大学官方网站，比如牛津大学、哈佛大学的官网。许多外部网站都会链接到大学网站上的研究资料及观点页面，这正是由于这些大学的研究成果受到了世界各地许多人的关注。

当牛津大学网络中的任何一个网站链接到外部网站时，该链接将提

升其所链接的外部网站的排名，这意味着牛津大学认可与该网站共享其巨大的声望。这也就是为什么我经常会被其他人要求，让我把在牛津数学系的网站链接到外部网站。这么做有助于提高外部网站的排名，毕竟能够在谷歌搜索排名登顶是每一个网站的终极“圣杯”。

谷歌的算法再强大，也不可避免地被那些了解数学原理的人用更加聪明的办法攻击并加以利用。在2018年夏天的某段时间里，如果你在谷歌上搜索“白痴”（idiot），首先弹出的便是唐纳德·特朗普（Donald Trump）的照片。一些激进分子知道怎样利用Reddit<sup>[1]</sup>在互联网的巨大影响力：他们在网上发布一个包含“白痴”这个关键词和特朗普照片的帖子让大家投票，两者的组合很快就登上了谷歌搜索的首位。但随着时间的推移，这些另类的搜索结果的峰值会通过算法本身而不是人工干预进行降级，并被算法慢慢剔除。谷歌不喜欢扮演“上帝”，但从长远来看，它相信的是数学的力量。

互联网就像一头会变身的妖兽，瞬息万变，每一纳秒就会有一个新网站诞生。随着网站的关闭和更新，新的链接关系也在不断地生成、变化，这意味着网站页面的搜索优先级需要动态调整。为了让谷歌跟上互联网不断发展的步伐，他们会使用一个有着相当可爱名字的工具，定期在互联网上搜罗网站信息并更新网站链接的统计点击量，这个工具叫作“谷歌蜘蛛侠”（Google spiders）。

技术控和体育教练发现，这种评估网络节点的方法也可以应用于其他“类网络”领域，尤其是足球。当评估对手时，评估对手的关键球员非常重要，因为他会控制整个球队的打法或者成为比赛的焦点。如果在比赛初期就能找到这名球员，并对他的发挥进行有效的遏制，那么在战略上就能冻结对手整个球队的战斗能力。

来自伦敦的两位数学家哈维尔·洛佩斯·佩纳（Javier López Pena）和雨果·杜塞特（Hugo Touchette）都是狂热的足球迷，他们决定研究一下，看看谷歌的算法是否有助于分析世界杯参赛球队的情况。他们是这样考虑的：如果把每位球员看作一个网站，一个球员给另一个球员传球就好比一个网站链接另一个网站，那么比赛中的传球路线就可以被视为一个网络；传球给队友是信任那个球员的标志，因为球员通常会避免传球给一个容易丢球的“笨”队友；一个不会积极跑动、有效控球的球员是很难拿到球的，所以只有有能力控球的人才会有人给他传球。

因此，他们决定使用国际足联在2010年世界杯期间提供的传球数据

来分析球员实力排名。当对英格兰队的比赛进行分析时，他们发现史蒂文·杰拉德（Steven Gerrard）和弗兰克·兰帕德（Frank Lampard）两位球员的数据明显高于其他人。这反映出一种情况：足球会频繁传给这两位中场球员，遏制他们在场上的发挥极有可能使英格兰队输球。最终，英格兰队在世界杯中确实没有走太远，复赛就被老对手德国队淘汰了。

以最终胜者西班牙队的数据来说，经过算法统计分析，西班牙队中没有明显的核心球员，这反映了整个球队很好地贯彻了“全攻全守”“快速短传”的战术思想，这最终促成了西班牙队走上冠军领奖台。

与美国许多依靠数据分析发展起来的体育项目不同，足球需要经过一段时间的积累才能利用数学和数据统计挖掘比赛背后隐藏的规律。但是到了2018年俄罗斯世界杯，许多球队都聘请了科学家在幕后提供技术支持——通过分析数据来了解对手的优势和弱点，这其中就包括对每支球队中传球依赖度的分析。

网络分析还应用于文学领域。安德鲁·贝弗里奇（Andrew Beveridge）和单杰（Jie Shan）用一套名为“网络科学”的方法分析了乔治·雷蒙德·理查德·马丁（George R.R.Martin）的史诗奇幻巨著《冰与火之歌》<sup>[2]</sup>（Song of Ice and Fire）。了解该故事剧情的人都知道，想要预测哪些角色会在剧本下一卷中出现不太容易，因为马丁先生会为了剧情需要，不惜“写死”哪怕是剧中最好的角色。

贝弗里奇和单杰决定在书中的人物之间建立一个网络。他们选定了剧中107个关键人物作为网络中的节点，然后根据关系重要程度为人物节点之间的连接线赋予权值。但算法如何评估节点间连接的重要性呢？该算法只是简单地计算剧情中两个人物名字在连续的15个单词内出现的次数。这并不是在衡量人物之间的友谊，而是在衡量他们之间的互动或联系频度。

他们选定这个系列的第三卷《冰雨的风暴》（A Storm of Sword）进行分析，因为剧情发展到这里也就基本稳定了。首先，他们对网络中的节点（或者说角色）进行了排名分析，三个角色很快脱颖而出，他们分别是提利昂（Tyrion）、琼恩·雪诺（Jon Snow）和珊莎·史塔克（Sansa Stark）。读过这本书或看过该系列电视作品的观众都不会对这个发现感到意外，但令人惊奇的是，一个不理解剧本内容的计算机算法也能分析出跟人相同的结论。这不是简单地计算某个角色名字出现的次数就能评估的，如果只是简单的这样做，会有其他人物的名字出现在排

名中。事实证明，算法对于这个剧情网络更加微妙的分析揭示出了剧中真正的主角。

随着剧情的发展，第三卷中的一些关键人物都被“写死”了，但这三个角色都在马丁的笔下幸存了下来。这就是一个优秀的算法存在的价值：从足球到《权力的游戏》，它在多种不同的应用场景中都能发挥作用。

[1] 一家社交新闻网站，其口号是“提前于新闻发声，来自互联网的声音”。——译者注



## 数学，幸福婚姻的秘诀

谢尔盖·布林和拉里·佩奇也许已经破解了引导你进入甚至连你自己都不知道自己要寻找的网站的秘法，但算法真的能帮助你找到灵魂伴侣吗？当你访问OKCupid<sup>[1]</sup>时，会看到一条带着些许骄傲情绪的标题——“我们用数学为你找到约会对象”。

这些婚恋交友网站使用“匹配算法”对个人资料进行检索，并根据个人的喜好、性格特点等进行配对。它们似乎做得不错！从某种程度上讲，这些算法选择的结果会比我们自己去相亲的结果更好：最近发表在《美国国家科学院院刊》（*Proceedings of the National Academy of Sciences*）的一项调查研究表明，以在2005~2012年结婚的1.9万人为研究对象，其中通过婚恋网站结识的伴侣相处得更幸福，婚姻也更稳定。

1962年，由两位数学家大卫·盖尔（David Gale）和劳埃德·沙普利（Lloyd Shapley）共同提出了史上第一个获得诺贝尔奖的算法——他们使用了一个匹配算法来解决“稳定婚姻问题”。遗憾的是，盖尔于2008年去世，没能参加颁奖典礼。沙普利在2012年与经济学家埃尔文·罗斯（Alvin Roth）共同获得了这个奖项。罗斯认为，这个算法不仅对人际关系问题的研究具有重要意义，对一些社会问题，包括医疗资源和学校招生名额分配等，也具有重大价值。

当时沙普利就被这个奖项逗乐了，他说：“我认为我是一个数学家，而这个奖是颁给经济学家的。”很显然，他对委员会的决定感到惊讶，他说：“我一生中从未上过经济学课程。”但是，他编写的数学算法已经对经济和社会产生了深远的影响。

沙普利和盖尔一起研究的稳定婚姻问题，感觉跟前沿经济理论没什么联系，更像是一个填字游戏。为了准确地描述该问题，我们假设有性取向正常的四位先生和四位女士，并按照他们的兴趣喜好对四名异性进行排序。该算法的难点在于如何给他们配对，并实现建立稳定婚姻关系的目的。稳定的婚姻关系意味着使所有的人获得较为满意的伴侣，不应该有任何一位成员因不满意算法分配的伴侣而选择在某个时刻离开，与其他人私奔。乍一看，即便只有四对关系，也很难安排得妥妥当当。

我们举个实例来看看盖尔和沙普利是如何利用系统和算法的方式来

保证稳定的配对关系。这四位先生分别用扑克牌中的K来表示，黑桃K、红桃K、梅花K和方块K；同样地，四位女士分别用Q来表示。每一位K和Q都列出了自己的偏好和习惯等参数。

对于K来说，选择方案如图4-5所示：

	K. ♠	K. ♥	K. ♦	K. ♣
首选方案	Q. ♠	Q. ♦	Q. ♣	Q. ♥
次选方案	Q. ♥	Q. ♣	Q. ♥	Q. ♥
三选方案	Q. ♥	Q. ♠	Q. ♦	Q. ♠
末选方案	Q. ♣	Q. ♥	Q. ♠	Q. ♣

图 4-5

对于Q来说，选择方案如图4-6所示：

	Q. ♠	Q. ♥	Q. ♦	Q. ♣
首选方案	K. ♥	K. ♣	K. ♠	K. ♥
次选方案	K. ♠	K. ♠	K. ♦	K. ♦
三选方案	K. ♦	K. ♥	K. ♥	K. ♠
末选方案	K. ♣	K. ♦	K. ♣	K. ♣

图 4-6

现在，假设提议每个K与同花色的Q配对。这肯定是不稳定的配对关系，为什么呢？梅花Q把梅花K列为她的末选对象，她和其他三个K在一起都会很开心。我们再来看看红桃K的列表：红桃Q是末选对象，方块Q是他的首选对象。在这种局面下，我们都可以想象到：某一日，风和日丽，梅花Q和红桃K私奔了。显然，同花色的配对关系不是稳定的婚姻方案。

我们该如何配对，才不会有私奔的状况出现呢？下面就是盖尔和沙普利所做的：利用多轮分析找到最终的稳定配对。第1轮中，Q都向其首选对象求婚：黑桃Q首选为红桃K，红桃Q首选为梅花K，方块Q首选为黑桃K，梅花Q首选为红桃K。似乎红桃K更受欢迎，有两个Q向其求婚。而红桃K选择他更青睐的梅花Q，所以拒绝了黑桃Q。因此，这一轮有三个待选和一个拒绝。

第1轮结果如图4-7所示：

K. ♠	K. ♥	K. ♦	K. ♣
Q. ♦	Q. ♠ Q. ♣		Q. ♥

图 4-7

被拒绝的Q必须放弃她的首选K，并在下一轮中向她的次选对象黑桃K求婚。这时，黑桃K有两个选择，第一轮中待选的方块Q以及新求婚的黑桃Q。对于黑桃K来说，他更偏爱黑桃Q，所以他会残忍地拒绝方块Q。

第2轮结果如图4-8所示：

K. ♠	K. ♥	K. ♦	K. ♣
Q. ♦ Q. ♠	Q. ♣		Q. ♥

图 4-8

接下来是第3轮。每一轮中被拒绝的Q都会向下一位K求婚，K们总是会选择相对更好的Q，所以这一轮，被拒绝的方块Q向方块K求婚（方块K一直孤单地等待，像是一个没有被选入足球队的孩子）。尽管方块Q在方块K的选项中排名很低，但他也没有更好的选择，因为其他三个Q更喜欢其他的K。

第3轮结果如图4-9所示：

K. ♠	K. ♥	K. ♦	K. ♣
Q. ♠	Q. ♣	Q. ♦	Q. ♥

图 4-9

我们用了一个可爱的Q与K配对的游戏来展现这个算法。最终，每个人都配对成功，所有的婚姻关系都很稳定，很圆满的大结局！这个算法目前在世界各地广泛使用：在丹麦用于小朋友匹配幼儿园；在匈牙利用于学生择校；在纽约用于给犹太教堂分配拉比<sup>[2]</sup>；在中国、德国和西班牙用于大学招生和学生择校；在英国被英国国家医疗服务体系（National Health Service）用于病人与器官捐赠配对，挽救了许多病人的生命。

在盖尔和沙普利研究的基础模型上，我们建立了婚恋交友网站用于配对分析的现代算法。当然，由于信息不完整，个人偏好会随时间、经历等因素而变化，实际情况中面临的问题会比这个复杂得多。从本质上讲，这些算法试图利用人们的偏好来进行匹配，从而形成稳定、幸福的婚配关系。有证据表明，这些算法很可能比人类的直觉更靠谱。

对于盖尔和沙普利开发的算法，你可能已经发现了一个有趣的不对称：让Q向K求婚。如果我们让K向Q求婚会怎样呢？相当引人注目的是，事实确实会跟我们想象的一样，即如果通过交换K和Q来应用该算法，将会得到一个完全不同的稳定配对。

方块Q最终会成为红桃K的伴侣，梅花Q会和方块K成为夫妇。这两个Q居然交换了伴侣，选择了优先级更低的对象。虽然两次不同的求婚都能形成稳定的配对关系，但Q向K求婚会给Q带来更好的结局。反之亦然，K去求婚，结果对K更有利。

这个算法在美国用于给医学院的毕业生分配工作。当毕业生意识到这样的配对方式不利于自己的就业时，发起了寻求公平对待的抗议活动。迫于各方压力，医院逆转了算法的选择方向，最终，毕业生得到了更好的有利于自己的分配结果。

这是一个重要的警示：我们的生活对算法的依赖越多，对算法的工作原理及运行过程的深入理解就变得越重要。否则，你很可能被算法“坑”了。

[1] 婚恋交友网站。——译者注

[2] 在犹太文化中，这是对有学识的人的尊称。——译者注

## 书商之战

算法的问题在于，有时会产生一些让人意想不到的结果。人类能察觉到怪异，而算法只知执行，它才不管结果是否谬以千里、荒诞不经。

我最喜欢举一个关于二手书商用算法来经营生意的例子。一位在加州大学伯克利分校（UC Berkeley）工作的博士后非常渴望拥有一本彼得·劳伦斯（Peter Lawrence）于1992年出版的经典著作《制造苍蝇》（The Making of A Fly）。发育生物学的专家经常会使用这本书，但到了2011年时，这本书已经停止发行很长时间了，所以这位博士后只能去找一本二手的。

他在亚马逊网上书城搜了一下这本书，二手书的价格大致在40美元，但他却发现有一个商家profnath对这本书开价1 730 045.91美元，并且还不包邮！还有一个商家bordeebook在做打折活动，但是价格更高，打完折要2 198 177.95美元，居然也不包邮（需要额外支付3.99美元运费）！

这位博士后把他看到的情况告诉了他的导师迈克尔·艾森（Michael Eisen）先生，导师也认为这两家的图书定价有些离谱，他觉得是有人（高学历的人）在开玩笑。但这两家书商在网上的评价还不错，应该不至于是不法商贩。profnath在过去一年中得到了8000多条推介，而bordeebook同期则得到了125 000多条推介。也许，这只是一个孤立的奇怪现象罢了。

第二天，艾森先生又到网上查看价格时发现，这两个书商竟然还在提价：profnath已经标到了2 194 443.04美元，而bordeebook则把价格提到了2 788 233.00美元。艾森先生决定花上几天时间，通过数据分析的科学手段来搞清楚这些奇怪的高昂价格背后的隐情。

最后，他发现了价格上涨背后的数学规律：用profnath当天的价格除以bordeebook前一天的价格总能得到0.998 30；用bordeebook当天的价格除以profnath当天的价格总能得到1.270 59。每个书商都为它们的网站设置了程序，使用特定的算法来确定售书价格。每天，profnath的算法都会检查bordeebook的图书价格，然后将其乘以0.998 30。这个算法非常有意义，表明profnath正在对网站中商品的价格进行算法控制，以削弱bordeebook价格的竞争力。bordeebook的算法也挺有特点，检测竞争



对手的价格变化情况，并将对手价格乘以1.270 59作为自己的售价。

表 4-1

	profnath 的价格 (美元)	bordeebook 的 价格 (美元)	profnath 当天的 价格 /bordeebook 前一天的价格	bordeebook 当天 的价格 /profnath 当天的价格
4 月 8 日	1 730 045.91	2 198 177.95	—	1.270 59
4 月 9 日	2 194 443.04	2 788 233.00	0.998 30	1.270 59
4 月 10 日	2 783 494.00	3 536 675.57	0.998 30	1.270 59
4 月 11 日	3 530 663.65	4 486 021.69	0.998 30	1.270 59
4 月 12 日	4 478 395.76	5 690 199.43	0.998 30	1.270 59
4 月 13 日	5 680 526.66	7 217 612.38	0.998 30	1.270 59

综上所述，价格每天都会乘以1.268 43（ $0.998\ 30 \times 1.270\ 59$ ），即价格将呈指数级增长。如果profnath采用了更激进的系数跟bordeebook打价格战，用不了多久，图书的价格不但难以维系上涨，反而可能会暴跌。

profnath所采用的算法已经描述清楚，但为什么bordeebook的算法会不断提高售书价格呢？没人愿意多花钱来买同一本书，或许在当初价格还不是很离谱的时候，它打算依靠自己良好的商家信誉和众多正面推介来招揽顾客。正如艾森先生在博客中所写的：“这种坐地起价的策略相当危险！它会让你的书在库房里接灰的，除非你根本就没有这本书……”

过了没多久，他就明白真相了。果然，bordeebook实际上就没有这本书！他们其实就是倒买倒卖而已。他们的算法还有其他一些功能，如查看其他书商在卖什么书，并把比自己价格低的书标记出来，利用差价买进，再依靠自己良好的商家信誉和众多正面推介以高价转手卖出。因此，该算法设定了涨价的系数1.270 59，这其中包含了购书成本、包装运费，以及他们的利润。

用对数函数可以轻松算出这本书大概在4月8日前的45天上市，售价约为40美元。这充分地显示出指数级增长的力量：短短一个半月，书价就飙升至数百万美元！4月18日，profnath的员工发现了该问题，此时该书的价格已经涨到了23 698 655.93美元（外加3.99美元的运费），他意识到这是个问题，立即采取了人工干预措施，把价格降至106.23美元。随后，bordeebook把这本书的价格改为134.97美元（ $106.23 \times 1.270$



59=134.97)。

《制造苍蝇》这本书的错误定价并没有给任何人带来重大影响，但算法的错误是可能导致重大后果的：如果将这样的算法用于股票期权定价将导致股市崩盘。算法可能导致意想不到的后果，是人们对先进技术存在恐惧的主要原因之一。试想一下，如果一家公司开发了一种算法，它的任务是最大限度回收碳资源，而众所周知，地球上所有生物体都是碳基的，它会理所当然地认为工厂里的人也是含碳生物，所以它在工厂里收集工人来生产碳资源是符合算法逻辑的。那么，可想而知，会发生什么？谁又能来阻止它？

算法是基于数学的，在某种程度上，它们就是数学运算。但是，它们并没有真正创造性地拓展应用它们的领域，所以在数学界，没有人会觉得自己受到了特别的威胁。我们真的不太相信算法会攻击它们的创造者，让我们失业。多年来，我一直认为，这些算法只是辅助我日常工作的工具——它们只是巴贝奇差分机的复杂升级版，主要还是在执行代数或数值运算，所带来的好处仅是将我从枯燥乏味的手写计算中解放出来而已。

直到几年前，人们依然认为人类能够理解他们的算法在做什么以及如何做的。就像阿达认为的那样，算法的输出不可能大于输入。但是，后来出现了一种新算法，这个算法可以在与数据交互时做出动态调整，运行一段时间后，它的程序员就可能弄不太明白他的算法为什么会做出这样或那样的选择了。程序带来了让人意想不到的收获，算法也变得更有创造力。这就是DeepMind团队所使用的新算法，这个算法使机器在围棋领域中战胜了人类，开创了机器学习的新纪元。

## 第5章 从“自上而下”到“自下而上”

艾伦·图灵

机器经常让我大吃一惊。

我第一次见到戴密斯·哈萨比斯是在他的算法于围棋对弈中获得大胜之前的几年，那时他参加了一个关于创新未来的会议。会议中，许多新创立的公司都在寻找风投和天使投资人。它们中的一些会改变未来，但大多数是“昙花一现”。风投和天使投资人的目的就是发掘最具潜力的投资对象。当时，哈萨比斯讲道，代码在未来可以自主学习、适应并进行自我改进，我听完立刻把他打发走了。因为我觉得，程序员通过编程教会计算机玩游戏，而计算机却会比教它的人玩得更好，这种输入寡而输出众的事情不大可能实现。我并不是唯一一个这么想的人。哈萨比斯也坦言，人工智能想要在短时间内得到认可并获得投资是极其困难的。

事实证明，哈萨比斯是千里马，而我却非伯乐。从最近的人工智能会议的标题就可以感受到哈萨比斯的观点所带来的变革性影响——“机器学习是新的42？”（42是道格拉斯·亚当斯（Douglas Adams）的小说《银河系漫游指南》（The Hitchhiker's Guide to the Galaxy）中“生命、宇宙以及任何事情的终极答案”。对于看科幻节目长大的人来说，这是非常熟悉的概念。）那么，是什么引发了这场新的人工智能革命呢？

答案很简单——数据。过去的5年间产生了现在世界上90%的数据。互联网每天创建的数据可达到1EB（ $10^{18}$ ），约为2.5亿张DVD光盘所存储的数据量。现在，人类两天内产生的数据量，就可以媲美从“文明曙光时代”到2003年所产生的数据量的总和。

海量的数据是机器学习进入新时代最重要的催化剂。在此之前，一直不具备任由算法在数据之中学习、遨游的客观条件。这就像是将一个孩子的感官全部隔绝起来，我们都知道，如果不与外界接触，他是无法获得语言及其他基本能力的。他的大脑已经准备好学习了，但没有足够的刺激或经验是不能正常发育的。

数据对于新的变革尤为重要，因此它也被称为“新的石油”。访问数据，相当于在21世纪的“油田”中进行开采。Facebook、Twitter、谷歌和

亚马逊备受瞩目，根源在于我们为其提供了免费的“石油储量”。严格意义上说，也不是完全免费，因为我们也在它们提供的服务中获得了数据的“交换”。比如，当我们在驾车使用Waze导航软件时，即在进行当前位置信息与到达目的地最有效的路线之间的数据交换。但是，很多人意识不到这种“交换”的重要性，不愿提供有价值的信息，因此其所得回报甚少。

机器学习的核心是在经验学习中改进算法的性能：对每个输出的结果进行自判断并即时调整算法公式，避免再次出错，在错误中不断学习进而得到改进。这就是数据访问十分重要的原因：数据越多，可用来进行训练的样本越多，得到的经验就越丰富，算法可以在不断迭代的过程中不断完善。程序员创建的元算法遇到不同的数据，就会发展成不同的新算法。

这种新方法的有效性同样震撼了整个人工智能领域，部分原因在于其底层技术并非独特的新技术。算法通过构建有解的问题层来实现，这种层模拟了人脑的工作原理，因此常被称为“神经网络”。在人类的大脑中，神经元之间通过突触相连，来自感官的刺激（如新出炉面包的诱人香味）会使得某一簇特定的神经元产生兴奋反应。当该刺激突破特定阈值时，次级神经元也将会被激活，由此会使人产生吃面包的冲动。比如，由10个神经元产生的兴奋反应可进一步激活次级神经元，但少于10个也许就不行。这种触发也可能取决于来自其他神经元输入的信号强度。

早在20世纪50年代，计算机科学家们就模拟该过程创造了感知器，其原理是：神经元就像一个逻辑门，接收输入的信息，然后通过计算来判断是否触发并产生兴奋反应。

如图5-1所示，感知器接收3个输入，每个输入的重要性不相同，用权重来表示：假设 $x_1$ 的权重是 $x_2$ 、 $x_3$ 的3倍，试计算 $3x_1 + x_2 + x_3$ ，然后将结果与某一阈值相比，高则触发感知器输出，低则不然。如果得不到预想的正确输出，将会重新调整输入的权重。例如， $x_3$ 对结果的影响可能比 $x_2$ 更重要，因此可将算式改为 $3x_1 + x_2 + 2x_3$ 。或者，可以用更简单的办法，微调一下阈值以直接触发感知器。我们还可以进一步对感知器进行改进，使其触发后的兴奋程度与函数超过阈值的程度成正比。此时，输出就可以作为对数据进行评估的一个标准了。

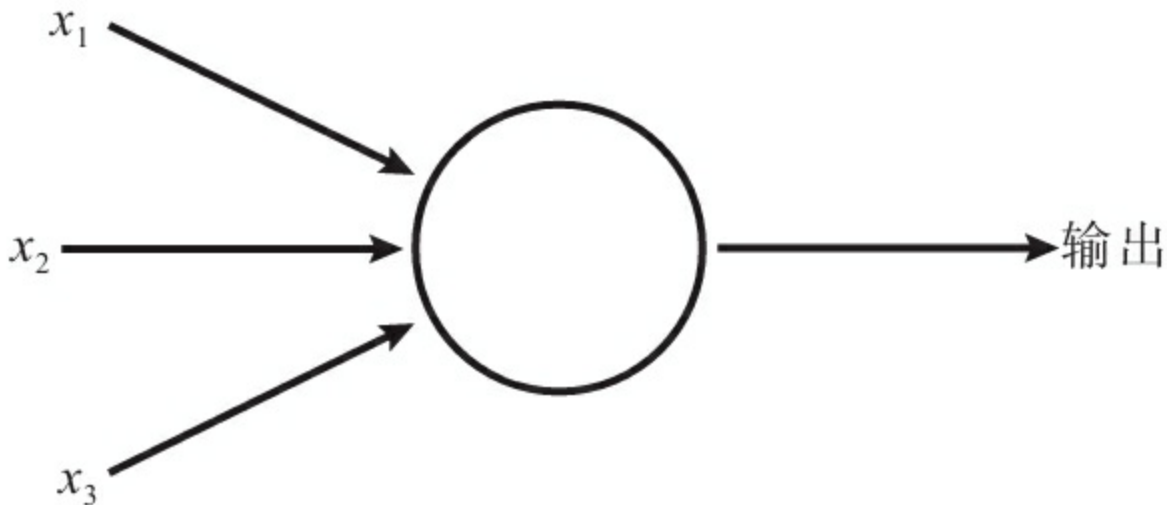


图 5-1

让我们模拟一个“确定今晚是否外出”的感知器，它做决定取决于三个要素：（1）电视上是否有好节目；（2）你的朋友是否一起出去；（3）今天是星期几。我们把它们看作三个变量 $x_1$ 、 $x_2$ 、 $x_3$ ，并按自己的偏好在0~10选择合适的数字来给这些选项赋值。比如，“星期一”赋值为1；“星期五”赋值为10。根据个人的偏好，某些变量或许会受到更多的重视。假如你是个电视迷，可能电视上播的无论是什么节目都会让你选择待在家里看电视，那么变量 $x_1$ 的值就会非常高。这个公式的作用就是通过调整权值和阈值来模拟你的行为。

就像大脑是由一连串神经元彼此联系、共同作用一样，感知器也可以通过分层来构造出节点间的级联效应，我们称之为“神经网络”，如图5-2所示。更为精细的感知器，如S型神经元，可进一步地平滑输出，而不仅限于“是”和“否”两种结果。

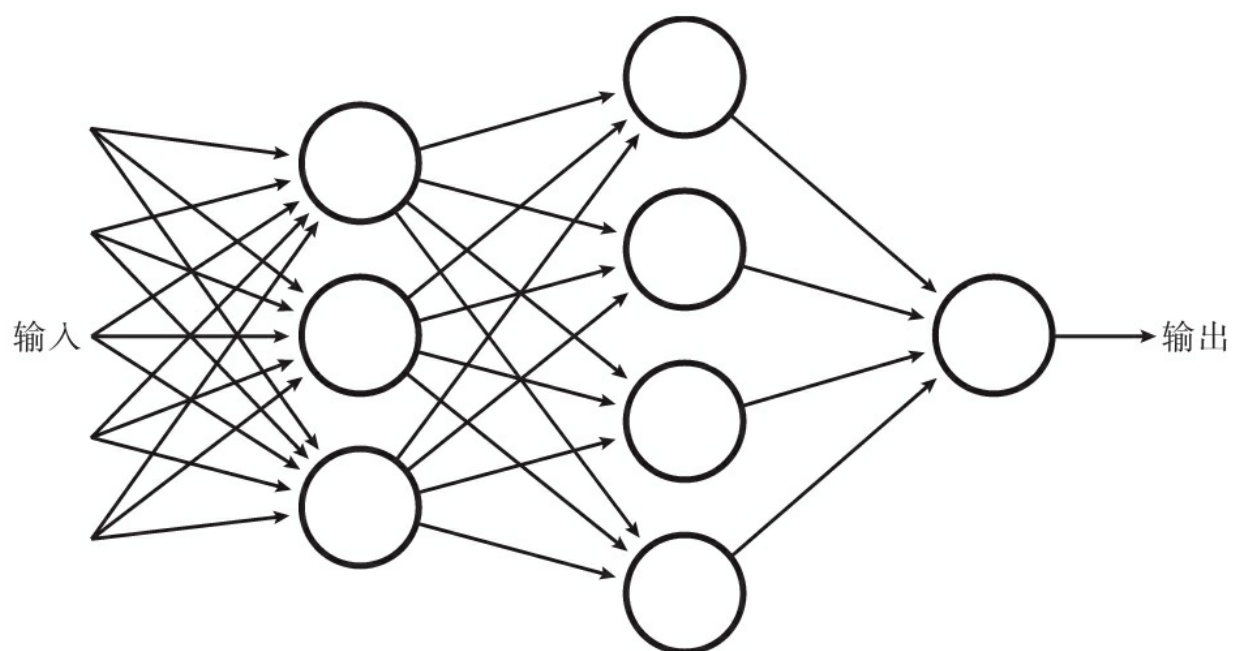


图 5-2

既然科学家们很早就懂得如何创建人工神经元，那么，从创建到使其高效工作的这个过程，何以耗费了那么长的时间？归根结底，还是因为数据，感知器的学习和进化需要数据的支撑。我们可通过编程为感知器分配初始的权值和阈值，通过真实的输入数据对其进行训练，每一次错误的预测都将会作为反馈，用以促进权值的调整和修正。这两个因素是创建有效算法应必备的。

## 看还是不看

人工智能首先需要突破的巨大难题是计算机的视觉问题。就在五年前，计算机识别物体还非常困难。而人类在这方面却有着压倒性的先天优势：人类可以迅速、准确地辨识出图片上的物体，并进一步对图片各区域进行分区域、分类识别。虽然计算机一瞬间可分析数百万个像素，但编写一个能将所有数据转化为有意义的信息的算法却相当困难。如何创建一个自上而下的算法来识别猫呢？虽然每幅图像中的像素构成是完全不同的，但人类大脑具备惊人的能力，可迅速处理数据并输出答案——“猫”。

人类大脑识别图像的能力广泛应用在为支付等领域构建附加的安全层（验证码）方面，以识别当前的交易对象是人类还是计算机程序，比如网上的抢票防作弊程序<sup>[1]</sup>。这其实就是一个反向的图灵测试。当看到一幅图像或一些特别的笔迹时，人类非常善于辨识出这幅图像或文字的特点，而计算机程序却不足以应对所有的变化。但是，机器学习使这一切发生了改变。

通过由猫的图像组成的数据库对算法进行训练，算法逐渐建立起一个问题层次结构，它可以向一幅具有高准确度的图像发问，并将其准确地识别为猫。这个算法与上一章中提到的略有不同，并且违反了我们所定义的构成一个好的算法的四个基本原则之一。算法在大部分时间都在工作，但达不到100%，所以关键是要让其尽可能达到“最”。对于业内人士来说，确定性算法到概率性算法的转变是一个重大的心理突破，类似于让数学家用工程师的思维方式去考虑问题。

既然如此，那当你在线购买最新的演唱会门票时，为什么还会被要求去识别一组图像以便进行人机区分呢？你所做的选择实际上是在帮助算法准备训练数据，这些数据随后会被载入算法，这样计算机就可以尝试学习你毫不费力就能做的事情了。算法需要借助带有标签的数据才能得到训练，我们真正在做的是训练视觉识别的算法。

用这些训练数据作为最好的问题类型可以训练算法，以便其区分猫和非猫的图像。每当算法识别错误，它将进行调整以便下次能够正确识别。调整的过程通常需要更改当前算法的参数或引入新的指标以求更准确地识别图像。这种更改不是通过程序员预判并自上而下实施的，而是



由算法自身通过高频次的数据交互自下而上构建的。

我在剑桥的微软实验室里，见识到了孩子们在家里使用的Xbox游戏机是如何识别人在摄像机前的动作的，并深深地被这种强大的自下而上的学习能力震撼。该算法可以很好地将人体的头、肘、手、脚区分开来。Xbox配有一款名为Kinect的深度感应摄像头，它可利用红外线技术记录障碍物与摄像头的距离：如果你站在客厅里的摄像头前，它会检测到它与你身体的距离比后侧墙壁更近，并且还能确定你身体的大致轮廓。

但人有男女老少、高矮胖瘦，特别是在玩游戏时还可能会摆出一些奇形怪状的姿势，所以该算法需要对人体的31处特定部位进行定位和识别。微软的算法可以达到对静态图像的认识（动态图像识别需要更强大的分析处理能力，强行为之会导致游戏卡顿、延迟）。

那么，算法是如何确定那些特定身体部位所对应的图像中的像素点呢？事实上，可通过一个简单的问题算法来实现，类似于“20个问题的猜字游戏”。第1个问题是：这个字位于词典的前半部分还是后半部分？答案如果是“前半部分”（词典此时被一分为二），则开始第2个问题：这个字位于“前半部分”的前半部还是后半部？此时词典被分为4个部分。依此类推，逐渐缩小范围。当提问到第20个问题之后，词典将会被划分为 $2^{20}$ 个不同的区域，总数达到100多万，远超过牛津词典中收录的词汇数目（约30万）。

如果我们想知道某一像素点属于身体的哪一部分，该怎样设计问题算法呢？在过去，我们都是通过苦思冥想或突发的灵感去设计一系列巧妙的问题。那是否能够通过计算机编程的手段，通过大量的数据交互与训练，选择、甄别出更好、更高效的问题集合呢？答案是“能”，它就是“机器学习”。

机器学习会从一些我们认为可能解决问题的候选问题开始，所以不算是零起点学习。学习的过程就是将最初的想法逐渐优化，从而形成有效的解决方案。那么，你认为什么样的问题可以帮助我们区分手臂和头呢？

假设被识别像素为 $x$ ，像素深度是其与摄像头之间的距离。若被测试人站在摄像头前，其身上的所有像素点的深度值（该点与摄像头之间的距离值）是大致相同的，而其身后墙壁上的像素点的深度值则明显偏

大。微软团队设计的问题策略是依次对 $x$ 周围的像素点进行询问：若 $x$ 位于头顶的位置，那么 $x$ 以北（上）的相邻像素点（可能是 $x$ 上方墙壁上的某一点）的深度值将远大于 $x$ 的深度值，而 $x$ 以南（下）的相邻像素点（可能是位于 $x$ 下方人脸某个部位的一点）的深度值则约等于 $x$ 的深度值；若 $x$ 位于水平伸直的手臂上，则手臂伸展方向上的相邻像素点的深度值约等于 $x$ 的深度值，而垂直于手臂方向上的相邻像素点的深度值则会由相对不变的状态迅速变大到与墙壁上的像素点的深度值相等。随着不断提问、回答、分析，该像素点位于身体哪个部位的答案将逐渐浮出水面。

这种不断提问的过程相当于构建了一棵决策树，每个后续问题形成树的一个分支。在初始状态下，选择任意的方向和深度阈值，如“方向向北，深度值小于 $y$ 时转到决策树的左分支，若大于则转到右分支”，依此类推。我们的目标是能够找到那些提供有价值的信息的问题。即使初始化状态是从随机问题开始的，一旦问题的数量累积到一定程度，如应用于10 000个带标记的图像时，我们就可以得到想要的答案。（比如说，我们判断出了在图像872中， $x$ 位于肘部；在图像3339中， $x$ 位于左脚。）我们想象有这样一些桶，每一只桶对应决策树的一个分支或者身体的某个部位，这样能够确认 $x$ 像素为肘部的所有的问题都可被放到其对应的一只桶中。当然，对于第一组随机问题集来说，确定它们所对应的桶还比较困难。但经过一段时间的学习或训练后，算法将逐步对角度和深度阈值进行优化，每个桶中的像素将得到更好的分类。

通过这个迭代的过程，算法会逐步调整阈值自我优化，以便于更好地识别像素点。重要的是，我们并不需要追求完美：如果桶中的1000幅图像有990幅均表明像素 $x$ 位于肘部，就意味着识别的正确率已然达到了99%。

算法自己找到了最优的问题集合，但程序员并不真正清楚它是如何得到这样的结果的。他们可以查看决策树中任何一个结点及其前后结点上所询问的问题，但树中的问题超过百万个，每个问题又略有不同，所以很难采用逆向工程法得出算法具体是怎样解决问题的。

试想一下，数量超过百万的问题如果全靠人工编程实现，即使再厉害、再勇敢的程序员也会闻之色变、望而生畏，然后望风而逃。但计算机却擅长做这种工作。编程团队需要有一些创造性的思维去相信通过对相邻像素点深度值的询问和比较，足以确定该点所在位置，而之后的创造则属于计算机了。

机器学习所面临的挑战之一是“过度拟合”<sup>[2]</sup>。针对训练数据提出用以区别图片的问题相对容易，但设计一个不针对特定数据的通用程序却非常困难，它需要从数据中抽象出广泛适用的规则。假设给定1000个人的姓名和护照号码，让你去设计问题集合对每个人加以识别，你可能会这样提问：“你的护照号码是834765489吗？好的，那你一定就是阿达·洛夫莱斯了！”但这样的问题只适用于现有的数据集，对于该组数据之外的新数据来说是难以实现身份识别的。因为护照号码是唯一的，不可能再有其他人拥有该组数据中的护照号码。

比如，根据给定的图表中的十个点，可得到一个方程，由该方程创建一条通过所有点的曲线。我们可以用一个十项式来表示这个方程，但这并不能很好地揭示数据中潜在的模式，而这对于理解和识别新数据是不利的。因此，有必要减少方程的项数，以避免出现过度拟合的问题。

过度拟合会误导你在建模时过分关注细节而错过数据呈现出的总体趋势，然后产生不符合客观规律的预测结果。如图5-3所示，这是一幅包含美国20世纪初人口总量的12个数据点的二维图。其总体趋势用二次方程式描述是最合适的，但如果我们选取的方程中 $x$ 项的幂超过2时会出现什么情况？例如，方程式中 $x$ 项的幂是从0变化到11，尽管由这个方程式确定的曲线与历史样本数据完美拟合，但一旦扩展到未来，曲线会突然向下倾斜，而这预示未来美国的人口数量会急转直下并且到2028年10月中旬就彻底消亡。这多少有些荒谬，或许是数学知道了一些不得了的事情吧！

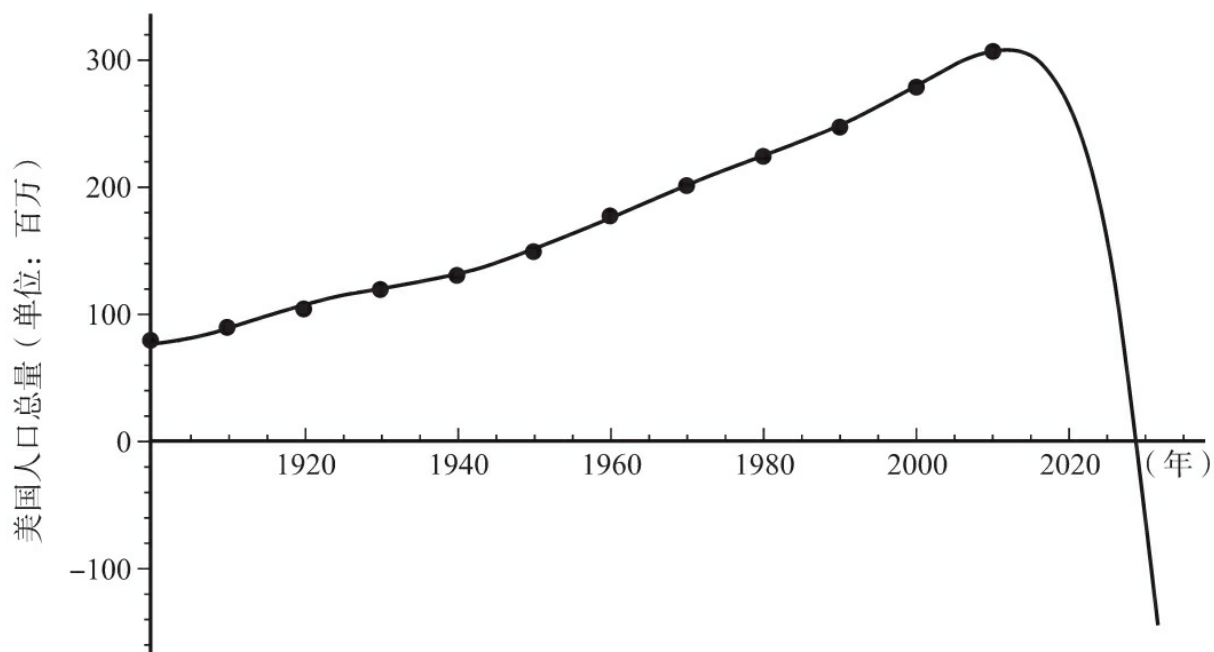


图 5-3

[1] 最贴切的例子是12306网站的购票确认验证。——译者注

[2] 指假设函数为了完美地拟合样本集，引入了过多的高次项。虽然这样做可以得到良好的样本拟合结果，但与实际情况完全偏离。——译者注

## 算法的幻觉

在过去的五年里，计算机视觉识别的发展进步让每个人都感到惊讶——新算法可以驾驭或识别的不仅仅是人类的身体。视觉识别能力的不足，一直是计算机不能与人类智能相媲美的最大障碍。比如，数码相机对图像细节的捕捉能力远超过人类大脑，但在图像识别的过程中，计算机读取到的是一个个孤立的像素点，而不是一副完整、有内容、有故事的图片。人类大脑在图像识别的过程中是如何处理数据的，至今仍然是一个未解之谜，更何况是我们的这些“硅朋友”<sup>[1]</sup>了。

接收到感官传递过来的信息后，人类大脑是如何将其当作一个整体看待的？例如，骰子的颜色（红色）和形状（立方体）在人类看来是相互关联的，大脑可以很好地将其融合成一个统一的整体。复制这种融合能力，一直是计算机解释图像所面临的挑战之一。按像素读取图像的方式，无法帮助计算机获得图像的整体信息。为了更直观地说明这一点，你可以拿出一张纸，在纸上面戳一个小洞，再把纸覆盖在一张A4大小的人脸图片上。仅仅通过移动带洞的纸，通过洞去观察，几乎是不可能分辨出照片中的人脸的。

五年之前的人工智能想突破这一挑战几乎是不可能的，因为当时机器学习还未出现，而程序员们一直以来都采用自上而下的思想设计图像识别算法。事实表明，想要通过“如果……那么……”这样的逻辑去识别从未见过的图像，是很难实现的。在自下而上的算法设计思路下，算法可以基于训练数据自主创建决策树，同时，网络上大量的训练数据——带标签的视觉数据（例如，Instagram上带评论的照片），也会为人工智能在视觉识别领域的迅速发展提供基本的支撑条件。

我们可以将图片上传至谷歌的视觉网站<sup>[2]</sup>来测试其算法的图像识别能力。去年，我上传了一幅我们家圣诞树的照片，经图像识别分析，算法认为图片上是一棵圣诞树的可能性为97%。这虽然还称不上惊天巨变，但已是非常了不起的进步了。乐观之余，算法的局限性也逐渐暴露出来，识别错误的也不乏其例。英国大都会警察局（British Metropolitan Police）在识别网上的儿童色情图片时发现，一旦图片涉及沙漠的景象，算法得出的结论就开始混乱不清了。

在最近的一次采访中，数字和电子取证部门的负责人马克·斯托克

斯（Mark Stokes）坦言：“由于沙漠的颜色跟人的肤色相近，而起伏的沙丘又极像裸露的身体曲线，所以很多人用作屏保的沙漠图片，有时就会被误认为是色情图片了。”

长久以来，计算机科学家们一直在努力开发一些特殊类型的图片，企图用它们“欺骗”人工智能。这些特殊的图片被称为“对抗图像”。它们实际上是利用了计算机看世界的不同方式，让人工智能的“眼睛”看到了并不存在的东西，我们可以理解为人工智能产生的“视觉幻象”。麻省理工学院LabSix团队<sup>[3]</sup>最新的研究试验表明，谷歌人工智能视觉识别系统对一只3D打印的海龟毫无辨识能力，无论以任何角度拿着这只海龟，甚至将它放置于大海等海龟应该在的环境里，其都会被识别为一支枪。这似乎点到了谷歌视觉识别的“完全死穴”。

该团队的做法是，在海龟的表面添加一层在人眼看来极像是海龟龟壳和皮肤颜色的纹理图案。但实际上，这种纹理是很巧妙地用由不断变化的来复枪图案组成的。即便是在变形干扰的情况下，计算机还是能将来复枪的图案识别出来，而且其被识别的可能性百分比远远超过海龟。尽管此前也出现过机器将猫的图片错误地识别为鳄梨色拉酱，但将图片稍微倾斜后，结果就恢复正常了。LabSix的贡献在于，无论从哪个角度看这只海龟，算法始终都会确信它看到的是一支步枪。

该团队把一幅狗的图像逐像素转换为两个在斜坡上滑雪的人，最终狗的图像在屏幕上完全消失了，但机器仍然将其识别为狗。尽管不知道图像是如何被算法识别的（算法完全像是一个“黑匣子”），但是他们还是设法去“欺骗”算法。

谷歌的研究人员创造了一种更适合算法口味的图像——“神奇贴纸”，它会导致算法忽略图片中的其他内容。其原理就是利用了算法会优先选择它认为对图像分类更重要的那部分像素。算法用于人脸识别时，它会自动过滤掉大部分背景像素，比如天空、草地、树木等。就是这一张小小的“贴纸”，一旦进入镜头，它旁边那根黄色的香蕉在机器的眼里就消失不见了。这种“贴纸”可以被标识为任意图像，比如一台烤面包机。无论算法正在识别的是什么图像，一旦“贴纸”进入机器的眼帘，它都会认为自己看到的就是烤面包机。这有点像一条狗完全被一只球分散了注意力，它的所见所想就只剩下了“球”，而其他的一切都从它的意识世界中消失不见了。先前的算法攻击测试绝大多数都需要事先获取被攻击对象（即被识别的图像）的大致内容，但这张“神奇贴纸”却什么都不用做——不管它试图干扰的图像里有什么，它都能正常工作。



人类不会轻易被这些黑客手段蒙骗，但这并不意味着我们可以彻底的免疫。比如，魔术师们就可以利用人类大脑的潜意识行为倾向，分散我们的注意力。这里有一个典型的例子——著名的两队传球：如果要求观众计算其中一支球队的传球次数，他们的注意力就会完全集中在球上，而看不到一个身穿着球衣的男子走到球员跟前，猛击他的胸部后离开了。可见，不仅算法存在视觉盲点，我们人类一样也有视觉盲点。

无人驾驶汽车所使用的也是视觉识别算法，很明显，这些算法也可能受到类似的攻击。想象一下，一个停车标志上贴着一张“神奇贴纸”，一个安检系统把枪错认为一只海龟.....将会发生什么？

我尝试在摄像头前做一些奇怪的、扭曲身体的动作，以对Kinect算法进行测试，看是否会对其识别结果产生影响。事实上，即便是训练数据中从未出现过的瑜伽动作，Kinect算法也能高度准确地识别出我身体的各个部位。由于缺乏外部刺激，即人身体能够做出的所有动作基本都是固定的，所以这个算法没有进一步“进化”。事实上，它也不需要进化，因为它正在有效地完成人类交给它的工作。而其他的一些算法可能需要不断适应新的变化，例如，算法若要给我们推荐喜欢看的电影、想读的书、想听的音乐，那么它必须是灵活的，足以应对我们不断变化的审美品位，并要能启发人类的思维和智慧去生成创造性的新潮流。

这就是算法能够不断学习、完善并适应新数据的力量所在。机器学习为其开辟了一个新篇章，即算法会像我们人类一样成长并走向成熟。

[1] 指计算机。——译者注

[2] 即Vision网站，网址为<https://cloud.google.com/vision/>。——译者注

[3] LabSix是一个独立运营的人工智能研究小组，由麻省理工学院（MIT）的本科生和毕业生组成。——译者注

## 第6章 算法的进化

卡尔·荣格（Carl Jung）

知识不仅建立在真理之上，也建立在错误之上。

现代算法是可以自学的，尤其是推荐系统算法，它可以根据每个人的喜好推荐有趣的东西给我们，比如说适合的电影、书籍、音乐，等等。算法通过与用户之间的交互过程，获取用户的个人偏好信息，并从中学习进一步完善自身，发现其中的关联关系，以便为下一位用户提供更优质的推荐信息。为了满足自己的好奇心，我研究了其中一种算法，想知道它到底能有多了解我。所以，在剑桥微软实验室测试Xbox游戏机所用的Kinect视觉识别算法时，我顺便拜访了一位同事，想向他了解一个实时推荐系统的算法原理。

该推荐系统的用户界面上随机排列了约200部影片，我要做的是如果看到自己喜欢的影片，就把它拖放到屏幕的右侧。其中的确有一些是我比较喜欢的影片。因为我是韦斯·安德森（Wes Anderson）的超级粉丝，所以就把《青春年少》拖到了屏幕右边。于是，屏幕中的影片位置自动调整，将我可能喜欢的电影放到靠右侧的位置，而将我可能不太喜欢的电影放到靠左侧的位置。仅仅通过一部电影是不足以训练算法的，所以大部分尚未被分类的影片仍然停留在屏幕中间的区域。

《王牌大贱谍》这部影片我觉得很无聊，属于我特别不喜欢的类型，所以我就把它拖到了屏幕左侧的回收站里。这一操作为算法提供了进一步学习所需的新信息，使屏幕中影片的位置又进行了一次调整，我似乎能感觉到算法对自己提供的建议很有信心：它将伍迪·艾伦

（Woody Allen）的《曼哈顿》推荐为我喜欢的电影。这部影片确实是我喜欢的，尽管算法是对的，但此时它还没有给我特别多的惊喜。它可能觉得我会比较喜欢《摇滚万岁》，所以将这部电影向右侧移动了。但事实恰好相反，我不能忍受这部影片，所以我把它拖到了回收站里。

算法本来以为我会喜欢《摇滚万岁》，但事实上我不喜欢，从这个过程中它获取到了重要的新信息。屏幕上的影片又一次进行了重新排列，并且此次调整的幅度非常大。这是因为系统后台的算法程序发生了微妙的改变——它根据我此次的选择学到了更多的“新知识”，并微调了

推荐系统的模型参数。它判断我可能喜欢《摇滚万岁》这部影片的概率过高，所以通过修正特定的参数来降低其值。虽然此前它从别的既喜欢韦斯·安德森又喜欢《曼哈顿》的那部分影迷处得知，他们也喜欢《摇滚万岁》这部电影，但这一条并不适用于我。

正是这种人机交互给算法提供了持续学习的新数据，使它可以不断进行自我调整以适配我们的喜好。在当今社会，这些算法在我们做出各种抉择时发挥了巨大作用：选择电影、音乐、书籍，甚至伴侣，等等。

## 如果你喜欢.....

电影推荐系统的算法原理比较简单。假定你喜欢电影A、B和C，而另一个用户也喜欢它们，但他还喜欢电影D，那么，D极有可能也是你所喜欢的。当然，现实中数据之间的逻辑关系并非如此简单。你喜欢电影A、B和C是因为这些影片里有你最喜欢的某位演员，但他并没有出演D这部电影。而另一个用户之所以喜欢A、B、C、D四部电影，是因为它们都是惊险刺激的间谍电影。

算法通过查看你所提供的信息，分析出你喜欢某类电影的原因，进而会把你和那些曾经做出过相同选择的人匹配、关联到一起。算法需要在大量的初始数据样本基础上展开工作，这一点跟许多机器学习算法是相同的。机器学习的一个重要特点是，人类必须参与到数据的分类过程中，以便让机器知道它所看到的到底是什么。这种管理数据的行为为算法提取潜在信息的模式做好了紧前准备。

算法在用户浏览影片库的行为过程中拾取关键特征值，如浪漫爱情喜剧、科幻片，或者是某位演员、某位导演的作品。但是，这种方法并不理想。首先，非常耗时；其次，分类的过程存在不客观因素，计算机最终学会的是已知的知识，而不能发现新的潜在趋势，从而导致计算机形成拟人态的思维定式。从最原始的数据中学习并发现模式是训练算法最好的方式。

奈飞公司开发出自己的电影推荐系统后，在2006年举办了奈飞大奖赛，期望通过竞争来发掘最优的算法。当时，奈飞公司已经积累了大量的电影评级数据，评分等级分为1~5星。于是，它公开了一个包含100 480 507个元素的电影评级训练集合，这些元素取自480 189个用户对17 770部电影的评价。然后，奈飞公司将17 770部电影的名称替换为数字序号，即变为匿名状态。比如，2666代表的可能是《银翼杀手》，也可能是《安妮·霍尔》，或其他任何一部影片。只有用户给这部电影的评分是已知的。

同时，奈飞公司还公布了一个包含2 817 131个元素的测试集合。测试集合的用户对电影所做的评价是未知的，因此参赛队提交的算法必须预测测试集合中所有的元素所对应的评价等级。比如，根据已有的数据预测出用户234654对2666这部影片的评价等级。重赏之下必有勇夫，奈

飞公司宣布设立100万美元奖金作为奖励，获奖条件是：以推荐效率提高10%的优势击败奈飞的自有算法。附加条件是：获胜者必须公开自己的算法并授予奈飞公司非排他性的许可，让奈飞公司有权使用这个算法向用户推荐电影。

除了100万美元的终极奖项，大赛还设立了几个进步奖：将上一年度成绩最好的推荐算法的效率提高至少1%的团队，将获得进步奖50 000美元。该奖项每年度都会有，但领取奖金的前提条件依然是需要公开算法的代码。

可能你会觉得从这样的数据里得不到有价值的信息，因为你甚至不知道2666所代表的影片是喜剧片还是科幻片。事实上，原始数据所蕴含的信息远比我们想象的要多。假设我们将每部电影视为一个维度，所有影片就构成了一个17 770维度的空间，那么每个用户就可以被看作这个17 770维空间中的一个点。每一部电影对应一个维度，用户对影片的评价越高，那么在该维度上此点偏离原点的距离就越远。当然，除非你是一个数学家，不然把用户看作17 770维空间中的点是很难想象的。实际上，我们可以把高维空间看作三维空间的扩展。假定只有3部影片被评级，我们可以用图形化的方式将用户与影片评级的关系表示出来。

假设电影1是《狮子王》，电影2是《闪灵》，电影3是《曼哈顿》。某一用户对这三部影片的评级分别为1星、4星和5星。用 $x$ 、 $y$ 、 $z$ 轴表示用户对电影1、电影2、电影3的喜爱等级，建立三维空间直角坐标系，如图6-1所示。这时，我们可以确定该用户在坐标系中的位置是 $(1, 4, 5)$ 。

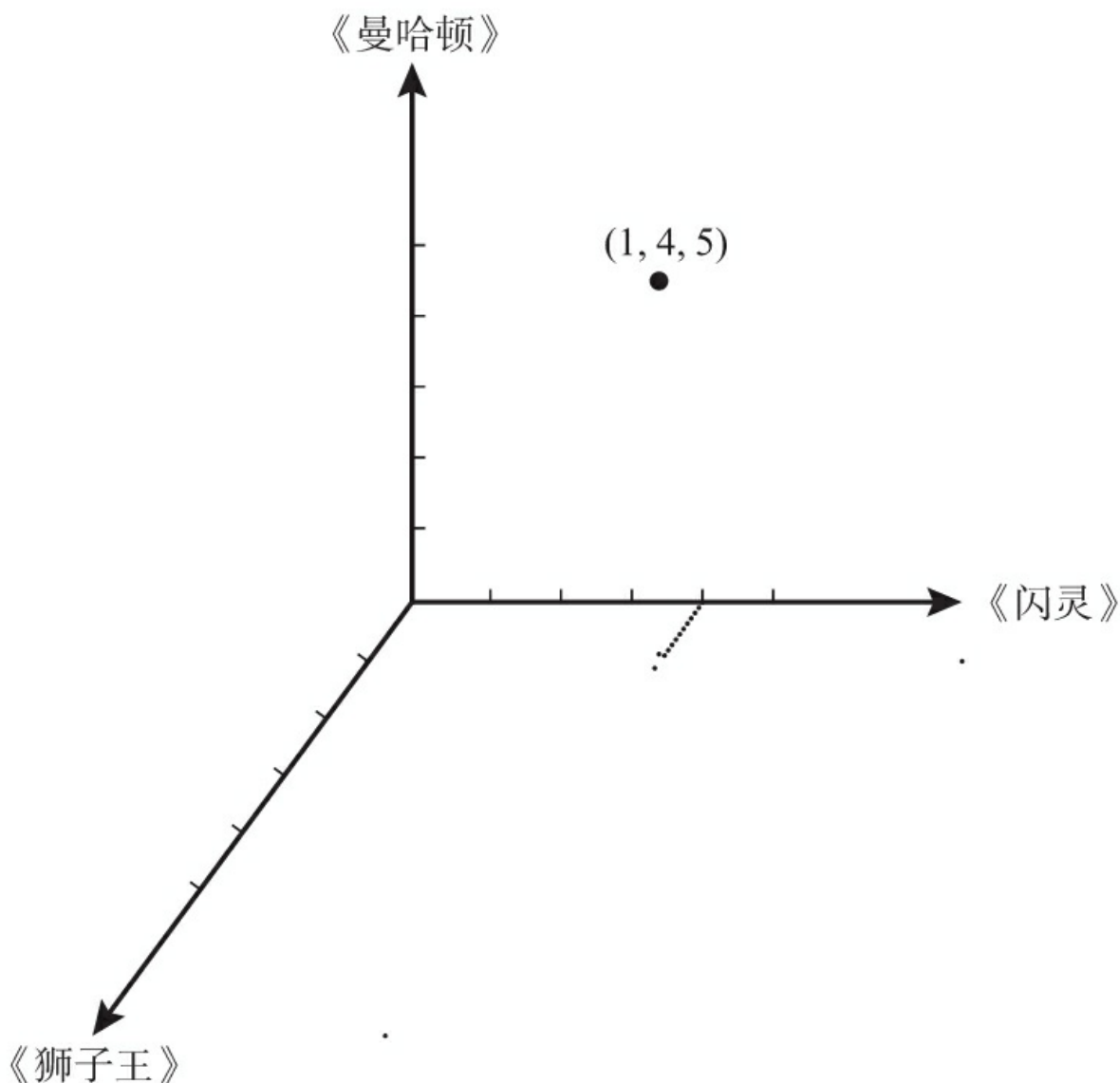


图 6-1

虽然在几何上无法绘制出17 770维空间以呈现用户在该空间上的所在位置，但数学可以。如果能把用户看成17 770维空间中的点，那么同样能把影片看作480 189维（用户数）空间中的点，此时，如果用户对影片评价越高，那么在该维度上此点偏离原点就越远。这些点分散在如此之大的维度中，很难发现其间存在的模式。因此，如果希望借助计算机找出数据中包含的信息，那么就需要降维处理。

这就好比一系列从不同角度得到的某人的头部剪影，其中一些更具代表性，更容易辨识一样。比如，希区柯克（Hitchcock）的侧影轮廓就比正面投影更易辨认。电影和用户就像脸上一个一个的点，以一个角度



投影，可能会看到这些点连成一条线，而以另外的一个角度投影，则可能并不会发现有明显的信息出现。

按照这个思路，我们或许能找到一种办法，将高维空间中的电影和用户对应的点同时投射到一个二维平面上，这样用户对应的点就会非常接近他喜爱的电影所对应的点。这种办法的巧妙之处就在于，能够寻找到揭示影片、用户所具有的潜在特征的合适投影。例如，图6-2是100个用户和500部电影匹配过后在二维平面中的投影，所使用的数据均来自奈飞公司的数据库。代表用户的点与代表影片的点很好地拟合，其余各处均未出现异常多余的点。我们可以通过这个投影找到数据中的信息。

通过比对点和其实际代表的电影，我们可以发现这个投影很好地反映出影片固有的一些特征：剧情片聚集在右上角，动作片聚集在左下角。

这就是最终在2009年赢得奈飞公司100万美元大奖的团队所使用的算法的基本思想。他们提取了有助于预测用户喜好的20个电影的独立特征，并将高维空间投射到这20个特征所构建的20维空间中。然后，借助计算机强大的运算能力，从海量的投影方案中挑选出最好的那一个。这正是计算机的强大之处，它的这种能力是人类的大脑和眼睛所无法企及的。

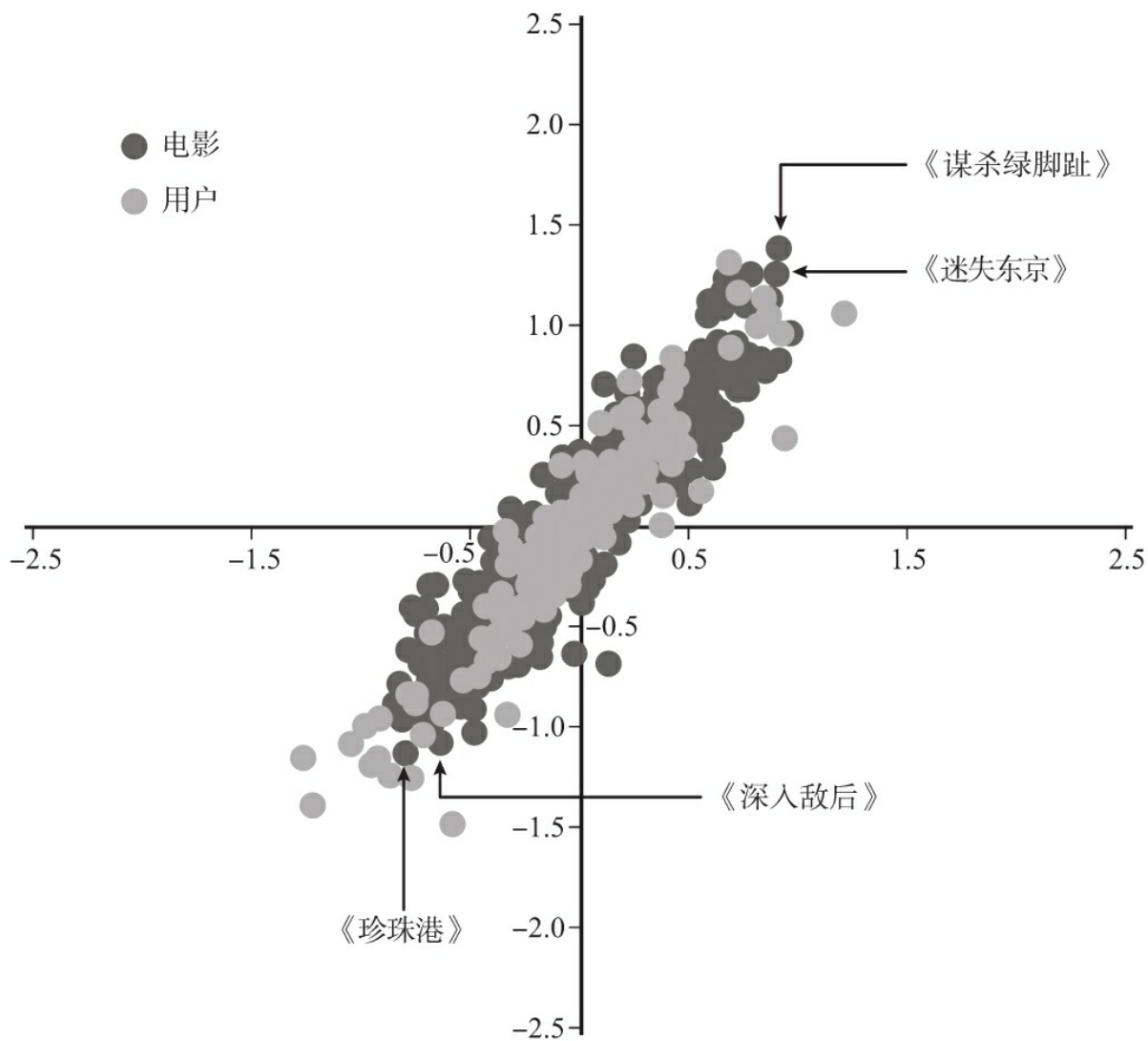


图 6-2

更有趣的是，模型挑选出的一些特征可以被明显地辨识出来，例如“动作片”或“剧情片”，而另一些特征虽没有明显的标识，但也都呈现出一种正态分布的趋向。

这正是我觉得计算机算法能给人惊喜的地方：它们有发掘新事物的潜力。从某种角度讲，深度学习算法提取出了人类无法用语言描述和表达的特征信息。就好比在没有建立颜色的概念，也没有红色或蓝色这种表示颜色的词汇的情况下，仅仅通过我们对所见事物表现出的好恶，计算机就能帮我们实现蓝色和红色的分类。因为存在太多的因素影响着我们的决定，我们无法准确地描述出为什么喜欢某部电影。决定个人偏好的人类代码遵循什么样的算法原理是隐性的，但计算机代码已经识别出

了引导我们偏好的特征，而这些特征我们仅可凭直觉感知，却无法表达出来。计算机代码在这一点上已经遥遥领先于人类了。

2009年6月，BellKor's Pragmatic Chaos团队向奈飞公司提交了他们研发的推荐算法，其推荐效率首次突破10%的要求并且达到了10.05%。按照比赛规则，奈飞公司宣布这场历时3年之久的比赛进入最后30天的决赛阶段。决赛阶段涉及的隐藏数据被分为两部分，一部分用来在公开测试中给每个团队打分，另一部分用来在隐藏测试中评判最后的赢家。如果没有其他队伍提交的算法超越BellKor's Pragmatic Chaos团队，那么他们无疑就是这场比赛的赢家。但是，就在决赛第29天的时候，Ensemble团队提交了他的算法，其推荐效率达到了10.09%，超过了BellKor's Pragmatic Chaos团队，位居公开测试排行榜的第一位。第二天，截至奈飞公司停止征集新算法的那一刻，两队的算法推荐效率又有了一些新的提升：BellKor's Pragmatic Chaos团队突破10.09%，Ensemble团队则非常接近于10.1%。但该结果并不能说明两队孰胜孰负，因此奈飞公司决定将大奖颁发给在隐藏测试中获胜的团队。可结果是两队的得分仍然相同，但由于BellKor's Pragmatic Chaos团队比Ensemble团队早20分钟提交了他的参赛算法，最终是他们带走了那100万美元。

鉴于第一次比赛的成功，奈飞公司希望能举办第二次比赛以激发更多的创新性想法，但是它遇到了一个难题：用户的个人隐私问题。奈飞公司曾在竞赛网站的页面上发出这样的公告：

训练数据集删除了所有的用户识别信息，只剩下评级等级和日期。这符合隐私权保护政策。即便知道自己所有的评级及日期，你也可能无法在数据中准确地识别它们，因为它们只是极小的一部分样本（总量不超过完整数据集的1/10），并且这些数据曾受到扰动。

那么，如果参赛选手知道用户所有的评级，这真的不算侵犯隐私吗？

得克萨斯大学奥斯汀分校的两名研究人员收集了这些数据，并将其与另一个网站——互联网电影数据库（Internet Movie Database，简称IMDb）上对电影进行评级的那些用户进行了比较，识别出了其中几名用户的身份。2009年12月17日，四名用户对奈飞公司提起诉讼，称该公司发布数据违反了《视频隐私保护法案》。其中一名用户表示，她是一个未出柜的同性恋母亲，有关她对电影喜好的数据可能会暴露出她的性取向。分析用户对电影的特殊喜好就有可能推断出其政治倾向或性取

向，这被称为“《断背山》因子”。最终，此案庭外和解，奈飞公司也因此取消了举办第二场比赛的计划。

数据像是一种新型“石油”，而我们却把它“倒”在互联网上了。谁拥有这些数据以及如何利用好这些数据，将是我们走向由这种新型“石油”推动的未来时，社会将要面临的一个重大问题。

## 如何训练算法

如果算法告诉你你可能会喜欢什么，那么就意味着你将永远都看不到算法认为你不喜欢的东西。知道了这一点，你是否感到一点隐隐的不安？对我来说，我很享受被引导找到自己喜欢的音乐的过程。此前，我经常不得不循环播放相同的歌曲，这也是为什么我总是听收音机的原因。现在，算法“连推带拉”地引导我从音乐库中挖到属于我自己的“宝石”。对于这些算法，我最初也曾有过怀疑：它会不会产生“口味”趋同的效应，导致所有用户最终都只关注音乐库中的一部分歌曲，而使另外一些歌曲失去听众呢？但后来我发现，这些算法都采用了非线性或混沌理论的数学思想：我和你喜欢的音乐风格只要略微不同，那么被推荐的歌曲将会大相径庭。

当在户外跑步时，我听了许多算法推荐的歌曲，它帮我找到了很多好听的新歌。几周前，妻子想要在她的生日聚会上跳舞，让我帮她挑选一些20世纪80年代的歌曲。我们花了几个晚上一起挑选她喜欢的音乐，并整理出一个精彩纷呈的歌单，聚会上所有的人都翩翩起舞。但生日聚会结束后的第二天，我又去户外跑步时发现，算法给我推荐的全部都是20世纪80年代的舞曲，于是我不停地按“下一曲”，但切换后出现的还是类似的歌曲。之后我通过听肖斯塔科维奇（Shostakovich）和梅西安（Messiaen）的音乐重新训练算法，花了好几周的时间才将这一切恢复如常。

电子邮件过滤器也是基于人机交互训练算法工作的。选用一些对你来说不是特别重要的邮件作为训练数据，将其中一部分标记为垃圾邮件，另一部分标记为正常邮件。通过分析这些邮件中出现的单词，算法开始构建垃圾邮件过滤的规则：带有“伟哥”这一词汇的邮件100%是垃圾邮件；带有“再融资”这一词汇的邮件99%是垃圾邮件；带有“火辣的俄罗斯人”这个组合词汇的邮件100%是垃圾邮件；带有“糖尿病”这个词的邮件不确定性比较大，因为似乎有一些人借助垃圾邮件四处传播治疗糖尿病的广告，但也有一些属于正常的邮件。所以，算法对这部分邮件进行了简单的统计分析，发现每20封含有“糖尿病”这个词的邮件中，只有1封属于正常邮件。因此，算法最终确定包含有“糖尿病”这一词汇的邮件有95%的可能是垃圾邮件。

电子邮件过滤器可设置不同的过滤级别。比如，只有在邮件有95%

的概率是垃圾邮件的情况下，该邮件才应该进入“垃圾邮件”文件夹。但现在更酷的是：虽然算法的训练数据是一组普通的电子邮件，但你的日常行为也将教会它识别你感兴趣的事情——算法会根据你所发出的邮件做出判断。假设你患有糖尿病，那么，根据你设置的最初过滤级别，所有带有“糖尿病”一词的邮件都会进入“垃圾邮件”文件夹。但渐渐地，随着你将越来越多的电子邮件（包括“糖尿病”一词）标记为“合法”，算法会重新校准已构建的邮件过滤规则，将这类邮件所对应的概率降至远低于95%的水平，这类电子邮件就会正常地进入收件箱而不是“垃圾邮件”文件夹。

不仅如此，算法还会自主创建新的算法，用于从所有包含“糖尿病”一词的邮件中区分出垃圾邮件和正常邮件。其方法是引入其他的关键词，例如“治愈”。机器学习算法将遍历你收到的每一封电子邮件，试图从中找出信息和关联，直到最后形成一个适合你个人生活方式的定制算法。

无人驾驶汽车也是基于这样的概率更新原理而设计的，虽然它的控制系统远比这复杂得多。该算法根据感知所获得的道路、车辆位置和障碍物信息等，来控制车辆的转向和速度。



## 偏见和盲点

奈飞公司的推荐系统算法可以提取出影片中那些人类都难以名状的特征，非常令人不可思议。这无疑挑战了洛夫莱斯的观点，即机器永远无法突破程序员思维的局限。现如今，机器掌握了人类所不具备的一项技能：对海量数据进行分析，并从中发掘出有价值的信息。

人类的大脑不擅长进行概率分析，这是进化的失败。概率直觉感知力的构建必须基于大量的试验，然后从中发现可能的趋势性。我们缺乏大量试验的机会，所以无法建立这种直觉。从某种程度上来说，机器代码的发展弥补了人类大脑在数据交互方面低速率的缺陷。因此，机器学习可以看作对人类思维方式的补充，而不仅仅是简单的复制。

概率是机器学习的核心。我们在第4章中讨论的许多算法实际上都是确定性的。这就好比一个人发现了问题的解决办法，然后通过编程促使计算机机械地在人所设定的规则下完成某一件事情。这就像牛顿的世界观：宇宙是由数学方程控制的，科学家的任务是发现其中蕴含的规律并用它们来预测未来。

20世纪的物理学家们向人们揭示了宇宙并不像我们此前想象的那样具有确定性。量子物理理论认为，世界就像上帝在玩骰子，结果具有不确定性，取决于事件发生的概率。概率思想主导下形成的算法具有非常强大的力量，这或许是为什么在新算法的世界中，那些受过物理思维训练的人比数学学者显得更游刃有余。这是理性主义与经验主义之间的对抗。但就我而言，经验主义占了上风。

那台机器是如何在不知道游戏规则的情况下仅靠一个可以左右移动的滑块，以及屏幕中的像素及分数变化情况就学会了雅达利开发的 **Breakout** 游戏的？算法计算的是当前状态下，左右移动滑块对得分情况的影响。由于移动所产生的影响是持续性的，可能是几秒钟时间，所以还需要计算延迟影响。这就非常难了，因为我们并不总是能够确定这之间存在什么样的因果关系。而这暴露了机器学习的一个缺点：它有时会把某些关联误认为是因果关系。动物们也会被这个问题所困扰。

有一个试验能很好地说明这一点，该试验拍摄了鸽子在笼子中的行为。试验人员在白天的某个时刻，将喂食机放进笼子。如果喂食机的门没有打开，鸽子们就会在旁边兴奋地等待着。此时，倘若在门打开之前

鸽子做了什么动作，第二天鸽子就会重复此动作。比如，鸽子看到门是关着的，它转了两个圈，然后门打开了，鸽子就会认为“转圈”是导致门打开的原因。它会觉得这是一种“奖赏”，所以下一次喂食机出现时，它就又会转两圈。<sup>[1]</sup>

美军用神经网络训练机器识别坦克图片的例子，是在机器学习社区里被反复提起的一个经典的反面教材。该算法的训练数据是一些带标签的图片（标记出哪些有坦克，哪些没有坦克）。通过对这些图片进行分析，算法获得了识别坦克的重要特征。在分析了数百张带标签的图片之后，研究人员用一些算法从未见过的图片进行测试，结果非常振奋人心，识别准确率达到了100%。

但将装有该算法的探测器部署到战场后，美军很快发现它毫无用处。令研究小组感到迷惑不解的是，当他们用美军所使用的图片做分析时，探测器所给出的识别结果竟然是随机的。直到后来通过深入研究他们才发现，只要图片拍摄于阴天，探测器就能做出正确的判断。

一想到训练数据，他们就明白问题出在哪儿了：研究小组只是获得批准有权使用坦克有限的几天，所以他们将坦克开到不同地方，拍摄了大量位于不同伪装位置的照片，但没注意到那些天一直都阴沉沉的。返回后，他们拍摄了一些没有坦克的乡村照片，但那时的天空是异常晴朗的。用这些照片作为训练数据，算法就误认为晴天和阴天也是区分坦克的重要特征。所以，一个坦克探测器就变成了一个对军方毫无用处的“阴天探测器”了。从这个事件中我们总结的经验教训是：机器是能够学习的，但前提是你要让它学习对的东西。

现如今，随着基于数据训练的算法广泛应用于社会的各行各业，如申请抵押贷款、治安决策、个人健康建议等，上面这个问题所造成的负面影响日益凸显。很多证据表明，算法中暗藏着歧视和偏见。麻省理工学院的一名研究生乔伊·布兰维尼（Joy Buolamwini）发现，她的脸相较于那些肤色较浅的同学，更加不易被所使用的机器人软件识别。当她带上一个万圣节用的白色面具时，技术识别很顺利，可一旦取下面具，她就在机器眼中“消失”了。

问题出在哪儿呢？该算法虽然针对大量的人脸图像进行了训练，但这些数据中黑色皮肤的面孔较少。本应公平公正的人工智能，从数据中学会了人类的偏见，这种偏差导致算法生成了许多令人不可接受也难以接受的结果：针对男性的声音进行训练的语音识别软件识别不了女性的

声音；某图像识别软件将黑人辨认为大猩猩；护照照片拍摄软件不适用于亚洲人，原因是它始终认为他们在拍照时闭眼了。硅谷的科技公司雇用的员工中有4/5都是男性白人，这正是布兰维尼创立算法正义联盟以对抗人工智能算法偏见的原因。

法律体系也面临着考验，申请抵押贷款、应聘工作、申领社会福利被算法程序拒绝后，人们有理由知道为什么。但是，由于算法是基于数据交互构建决策树的，证明其决策的合理性并不容易。

虽然有些人主张采取法律措施补救，但执行起来非常困难。2018年5月生效的欧盟《一般数据保护条例》的第22条规定：数据主体有权反对“完全依靠自动化处理做出的决策”。若个人对自动化决定不满，有权主张人工介入，以表达自己的观点并提出质疑。对于计算机所做出的任何决策，数据主体有权获得“有关所涉及逻辑推断的有意义的信息”。关于这一点，我只能表达我的个人观点：祝你好运！

人工智能领域一直呼吁开发一种元语言，机器可用它来证明自己的决策是合理的。而在此之前，我们必须谨慎对待这些算法对日常生活的影响。绝大多数算法都有其专长，不善于处理无规律的行为，当意料之外的事情出现时，算法只能选择将其忽略，而人类对这类场景却能表现出非凡的应变能力。

没有免费午餐的定理表明，在任何情境下都能做出准确预测的通用性算法是不存在的。机器学习的目标不是建立放之四海而皆准的通用模型，而是构建关于特定问题有针对性的解决方案。这个定理还表明，即使只向算法出示一半数据，它还是有可能将未出示的另一半数据伪造出来，以保证对它自己所进行的数据训练的完整性。但当需要分析的数据属于另一半未出示的数据时，它就会决策失准或失败。

数据本身永远无法自给自足，它必须与知识相结合。正是在这一点上，人类的思维和智慧似乎能更好地应对环境的变化并对全局进行把控——至少在目前看是这样。

[1] 原理同巴甫洛夫的条件反射试验。——译者注

## 机器之间的战争

是改变和适应新挑战的能力，让AlphaGo得以诞生。谷歌的DeepMind团队用一段时间的“监督学习”构建了他们的算法，这就好比一个成年人帮助孩子学习成年人已经掌握的技能。作为一个物种，人类之所以取得进步，是因为我们积累了知识，并以比最初获得知识时更有效的方式将知识传递了下去。作为一个数学家，我也是用上大学的几年时间快速学习了前辈们近几个世纪发现的数学理论，而不是凭借一己之力去重新发现所有的数学知识，以此站到学科前沿的。

AlphaGo也经历了同样的学习阶段——我们称之为基础学习阶段。互联网上有数以百万计的棋局，其中不乏高手之间的对弈。这是一个极佳的资源宝库，通过检索就可以找到给对手致命一击的决胜杀招。这样一个庞大的数据库使得计算机能够建立一个概率的概念，即给定一个特定的棋盘位置，算法可分析出每一步落子对获胜的影响概率。仅考虑每一盘棋已有走法的优劣是不够的，因为未来的对手可能不会使用数据库中失败棋手所用过的棋路。这个学习阶段为AlphaGo下围棋提供了必要的基础准备，但仅仅使用这个数据库还是不够的。

第二阶段被称为强化学习阶段。从长远来看，它为算法自我的发展确立了优势，提供了可能。算法通过自我对战、强化学习来提高对弈水平，即与之前的“自己”不间断地训练以提高下棋的水平。如果某些有望获胜的棋招失败了，算法就会修改这些棋招的概率。这种强化学习会综合生成大量的新数据，有助于算法发现自身可能存在的弱点。

局部最优是这种强化学习的弱点之一。机器学习有点像攀登珠穆朗玛峰，如果目标是爬上这座世界上最高的山峰，但是此时你既不知道自己身处何处，又被蒙上了眼睛，那么可选择的策略就是以你所在位置为起始点，如果下一步能够爬得更高，那么就继续往上爬，如此往复，直至抵达顶峰。按照该策略，你可以抵达所在地理区域内（当前环境下）的最高点。一旦越过这个顶点，高度就会下降，你就有可能滚落下来。但这并不意味着，在不断努力之下，你会抵达山谷对面另一座更高的山峰。这个高点是相对的，在数学中被称为“局部极大值”。就好比你好不容易爬上一座山的山顶，却发现它不过是在高耸的群山环绕下的一座小山包。如果AlphaGo训练的算法只能在局部极大值的情况下击败对手，那终极结果会是怎么样的呢？



在与李世石对战的前几天，欧洲冠军樊麾帮AlphaGo进行赛前集训时发现了它的一个弱点，似乎就是这种情况。这或许说明人类的下棋数据将算法导向了局部最优，而实际更优或者最优的下法与人类的下法存在一些本质的不同，即人类在事实上“误导”了AlphaGo。算法很快就学会了如何重新评估自己的落子，以最大限度地提高再次获胜的概率。是新对手把算法“逼下山”，促使它找到了再攀高峰的新方法。

DeepMind团队目前又开发出了新一代的AlphaGo—AlphaZero，它打败了曾书写历史的各版本AlphaGo前辈。这个名字的由来是：由于是通用棋类人工智能，因此去掉了代表围棋的英文“Go”；没有使用人类的知识，从零开始训练，所以用“Zero”；两者相结合就得到了“AlphaZero”。它已不再学习人类的棋谱、走法，而是完全依靠自我对弈来迅速地提高棋艺，从而走出人类对围棋认知的局限与定式。就像雅达利的游戏算法一样，给定棋盘上361个（19×19）交叉点以及它们的得分规则，然后在自我对弈中试验棋步。沿用最初在构建AlphaGo时所使用的强化学习策略，由“白板”<sup>[1]</sup>状态开始“自学成才”是AlphaZero的独门秘籍。DeepMind团队甚至也为新算法呈现出来的强大能力而感到震惊：它已不再受人类的思维和游戏方式的限制了。

AlphaZero自我训练的时间仅为3天，完成的自我对弈棋局数量就达到了490万盘。人类花3000年才能实现的，它却只用了3天。在对阵曾赢下李世石那一版的AlphaGo时，AlphaZero取得了100：0的压倒性战绩。经过40天的训练之后，它就所向披靡了。它甚至能够在8小时内学会如何下国际象棋和日本将棋，水平甚至超过了市面上两个最好的国际象棋程序。

AlphaGo项目的负责人大卫·西尔弗博士解释了这种“白板”学习在多个领域的影响：

如果能够实现“白板”学习，就相当于拥有一个可以从围棋移植到其他任何领域的桥梁。这种算法是通用的，它会将你从所在领域的细节中解放出来，它普遍适合于任何领域。AlphaGo并不是要打败人类，而是要发现做科学研究意味着什么，让程序能通过自学最终学习到哪些知识。

DeepMind的口号是：首先解决智力问题，然后用它来解决其他问题。他们确信未来已在路上。但这项技术能走多远？它在创造力方面能与最优秀的数学家相媲美吗？能绘画或是创作音乐吗？能破解人类大脑

的秘密吗？

[1] 指所有知识均由感官和经验而来，即从零开始的学习。——译者注



## 第7章 数字绘画

汤姆·斯托帕德（Tom Stoppard）

事物的可预测性与不可预测性构成了我们熟识的世界。

几年前的一个周六下午，我参观了德国著名艺术家格哈德·里希特（Gerhard Richter）的画展，当我在展厅的蛇形画廊中前行时，这些作品的色彩和线条带给我和同行者们深深的震撼。艺术作品就是通过各种形式给人带来美的感受，从而使人们获得精神上的愉悦与放松。

当时展出的是里希特的《4900种色彩》系列绘画作品。在去往画展的路上，我的夫人跟我介绍说：“他是这个地球上硕果仅存的尚且健在的最受敬重的艺术家！”在来画展之前，因为绝大部分时间都沉浸在抽象的数学世界中，我对格哈德·里希特和他的作品并不熟知，更遑论对这些有极大的兴趣了。正因为如此，我的夫人对我各方面知识的匮乏感到惊讶和无力。对于画展，我并未抱以很高的期望，但令人意外的是，里希特的作品竟然能够以一种有趣的方式跟我夜以继日沉迷着的数学抽象世界直接对话。

《4900种色彩》系列绘画由196块画板组成，为了叙述方便，我们将其称为“单元”。每个单元都由25块边长为9.7厘米的正方形色块以5行5列的形式组合而成，每一个色块都是一种精心挑选的颜色。整个作品由4900个色块组成，这正是画作标题的出处。这196个单元可自由排列组合，通过这样的排列组合，我们能得到不同版本的绘画作品，这也就使我们可以收获不同的视觉展示效果。我参观的是该系列的2.0版，196个单元按照4个一组划分，构成10行10列共100个色块的画作，整个展厅共陈列49幅。

这些“像素化”的画作激发了人类最本能的冲动，去探索其中的奥妙。对于我而言，我的注意力被一幅3个黄色色块并排连续出现的画作吸引了过去。在历史的长河里人类能够生存下来并发展至今，全都有赖于我们善于探索混沌的世界，从中发现规律并运用这些规律，这是我们能够不被远古时期潜行在灌木丛中的猛兽吃掉的根本原因。3个并排的黄色色块意味着什么？也许什么都不是，也许是“草丛中饥肠辘辘的狮子”。

诸多的心理学家，如荣格、罗夏（Rorschach），还有马特·布兰科（Matte Blanco）等，都认为可以通过图像的特征，如形状、颜色、对称性等探究人类的心理。这些心理学家也都在他们的实践过程中做着这样的探索和实践，例如荣格就曾让他的病人画曼陀罗，而罗夏的墨迹测验是最著名的投射法人格测验。

我们的大脑总是处于发现、分析并设法解决问题的状态，数学家们对模式（这里可理解为典型案例）是非常敏感的。当我游走在蛇形画廊中，满眼都是这些“像素化”的画作时，心里开始琢磨起来：在这一幅幅画作里会不会暗含着一些超出其本身表达的有趣信息呢？

我发现，无论是水平方向还是垂直方向，两个相同颜色的色块连续出现的次数，总是略超过三个或四个相同颜色的色块连续出现的次数。于是我坐下来，开始琢磨里希特对颜色的选择是不是一种随机行为？因为，随机性的聚集现象会以人们时常见得到，但是不太容易联想到的方式出现。例如，虽然公交车在始发站都是按时刻表准时发车的，但是当我们在公交站牌前等车的时候，经常会碰到三辆同一路公交车一起到站的情况，这就是由路程中的不确定性问题导致的随机性的聚集现象。

我猜测3个黄色方块连续出现正是这种随机性的聚集现象产生的结果。于是，我做了一些计算来验证我的猜测是否正确。我们将计算过程简化一下，先讨论横向色块的分布。假定可选颜色有25种，且每次选择是随机的，计算每行中两个同色色块一起出现的概率时，采用逆推法，即先计算不同时出现的概率。

假定第一个色块为红色，那么下一个色块不是红色的概率是 $24/25$ ，第三个色块不是第二个色块所选颜色的概率也是 $24/25$ 。依此类推，若每行包含10个色块，那么两两颜色不同的概率为：

$$\left(\frac{24}{25}\right)^9 = 0.69$$

所以，两个同色色块一起出现的概率就是0.31。纵向上的推导也可以采用同样的方法。这意味着在一个10行10列的画布上，同色色块两两（行或列）连续出现的次数约为3次。这一结果在当天展出的画作上得到了很好的验证。

按照同样的推导方法，整个系列的4900个色块之中，3个同色色块连续出现（水平和垂直方向）的次数约为6次，而展出画布上3个同色色

块并排出现的次数略超过6次，这就是随机性产生了聚集现象。由于推导随机性本身就要考虑不确定因素带来的影响，而正因为不确定性因素的存在，所以导致了计算结果不可能达到完全精确。随机性的推导本身就不是一门精确的科学。

在参观完画展之后，为了验证我的猜测和计算的结果，我做了一些调查，得到的证据进一步证实了我的猜测。果然，里希特是随机选取颜色的。他将25个大小、形状相同但颜色不同的方块放入一个袋子，然后从袋子中随机抽取一个方块以确定绘制的颜色，就跟抽奖一样。用这样的方法，里希特绘制出了《4900种色彩》系列的196个5行5列的单元。事实上，他总共可以做2525个不重样的单元。 $25^{25}$ 有多大？它是一个36位数，将这么多的单元首尾相连，长度可达 $4.3 \times 10^{31}$ 公里（约 $4.5 \times 10^{10}$ 亿光年），远超过我们目之可及的范围（哈勃望远镜的观测范围大约为270亿光年）。

在参观完这次画展之后的几天里，我无可救药地沉迷于这些五颜六色的方块之中而不可自拔，我想我的夫人再也不会带我去参观类似的画展了。我当时所参观的只是《4900种色彩》系列的2.0版，该系列的1.0版是把所有的单元组合在一起，形成一个巨大的70行70列的布满色块的画作。既然已经有了1.0版、2.0版，那么196个单元随机组合到底可以形成多少种展示版本呢？答案可由某个方程求解，这个方程曾引起17世纪著名的数学家皮埃尔·德·费马（Pierre de Fermat）的关注。

我难以自抑地将我的发现告诉了策展人汉斯-乌尔里希·奥布里斯特（Hans-Ulrich Obrist）。过了一段时间，我收到了里希特的来信，在信中，里希特询问我是否同意将我的观点译为德文与他的作品一并出版。他觉得十分不可思议，并幽默地表示，不知道未来还会有多少个数学公式从他的作品里面“冒出泡来”？

科隆大教堂的彩色花窗也是里希特的设计作品之一。其创作过程与《4900种色彩》类似，只是添加了一些对称性的设计要素。里希特对随机产生的三种设计进行反复比对并进行排列组合，最终确定了六扇窗户的色彩设计。虽然对称性并不十分明显，但与罗夏墨迹测验<sup>[1]</sup>比起来，它更容易让我们产生共鸣，更具亲和力。

从某种程度上讲，里希特是在使用编码的方式创作。袋子里方块上的颜色是艺术家选择的，但是每次绘制哪种颜色是不受艺术家控制的，

取决于随机性。艺术家掌控的是架构的设计，而不是具体的执行过程。

这种允许局部过程自主发展的方法，是早期设计创造性算法的基本策略之一，其生成的代码常常使程序员感到惊讶。如何借助程序员最初写的简单算法，让其自主发展出新的、有价值的、出乎意料的算法呢？答案就是像里希特所做的那样，让确定性与随机性相结合，这个策略也是攻破洛夫莱斯测试的潜在方法。

[1] 由瑞士精神科医生、精神病学家罗夏创立的人格测验，是少有的投射型人格测试。——译者注

## 什么是艺术

什么是艺术？说到这里，我就需要问一些前置问题了。

为什么会有人想用电脑来创造艺术？

动机是什么？

艺术不应该是一种情感的流露吗？

艺术不应该是打上人类烙印的作品吗？

为什么要让电脑生成艺术？

艺术是商业吗？创作者们是想通过按下“打印键”来赚钱吗？

电脑是扩展我们创造力的新工具吗？

为什么我们人类要创造艺术？

为什么里希特的作品被认为是艺术，而一本油漆试色谱不是？

我们真的了解我们所谓的艺术到底是什么吗？

这一切从何而来？

尽管人类起源于距今约600万年的非洲，但直到260万年前（旧石器时代）能人制作出用于切割的打制石器工具时，人类才真正体现出这个种群所独具的创造力。但这一创造并没有引发艺术创作的热潮。根据南非布隆伯斯洞穴（Blombos Cave）的考古发现，那时的人类拥有贝壳串珠、雕刻过的赭石和处理赭石的工具组，这表明大约在10万年前，人类就有了对艺术品创作的热情。赭石是氧化物类矿物刚玉族赤铁矿，主要含三氧化二铁，这是制造赭红颜料的原料。那时候已经有了制作颜料的工具。这些颜料作何用途，目前考古学家还未有定论，他们设想了以下几种可能：涂在身上或是绘制在皮革或其他物品上作为装饰，抑或用于墙壁绘画。在南非的这些洞穴中，考古学家们并没有发现确切完整的线索，毕竟环境的因素使这些地方的岩石壁画存留至今是非常困难的。

但是，考古学家们在地球其他地方的一些位于地下深处的洞穴岩壁上，发现了一些人类早期绘制的图案，其中人手轮廓的图案频繁出现。在印度尼西亚东部苏拉威西岛的马洛斯石灰岩洞穴的内壁上，有很多红色的人手轮廓的喷绘印记。这些壁画约绘制于4万年前。我们有理由相信，绘制人手轮廓喷绘的这位艺术家是这样做的：将赭红颜料含在嘴里，然后以手（自己的或是模特的）为模板覆盖在岩壁上，而后将口中的赭红颜料喷涂在手以及附近的岩壁上。

洞穴壁画被视为史前人类艺术活动和抽象思考的标志，手印即一种“存在”的表达。正如雅各布·布洛诺夫斯基（Jacob Bronowski）在他最著名的BBC纪录片《人之上升》（The Ascent of Man）中所说的，手印就代表着“这是我的标记，我是人类”。

除了手印，考古学家们还在该壁画中找到了一些人物及当地特有的有蹄动物的图案。其中，类似猪的当地动物的图案已被证明至少有3.54万年的历史了，这是至今世上所知的最古老的具有丰富修饰的绘画作品。考古学家们检查了洞穴里的矿物层（主要是方解石），其厚度不到10厘米，有些覆盖在绘画上，有些则是把绘画夹在中间。分析测定方解石，就能确定这些矿物质是何时被水带到洞壁上的，也就是说确认这些沉积物的年代就能更加精确地确认绘画的创作时间。

因此，洞穴中最古老的手印的推测年代只是覆盖在手印上的矿物质的最低年龄而已，手印本身有可能比估算的时间还要早好几千年。

4万年前似乎发生了什么，平地一声雷，开启了人类这个物种在艺术创作上的新纪元。

考古学家们在西班牙的洞穴中发现的人类手印壁画可追溯到4.5万年前，该洞穴壁画有可能由更早的尼安德特人创造。<sup>[1]</sup>随着智人由非洲迁徙抵达欧洲，大约5000年后，尼安德特人神秘地消失了。<sup>[2]</sup>在最近一次对西班牙洞穴某些岩石壁画的年代鉴定中，考古学家们发现其形成时间可追溯到6.5万年前。当时，智人并未抵达欧洲，因此，可以肯定创作这些壁画的另有其人。那么，谁才是最早的艺术家呢？在印尼爪哇岛发现的贝壳雕刻图案制作于50万年前，这显然是另一种早期人类——直立人（homo erectus）的作品。所以，如果有人可以被称作最早的艺术家的话，那既不是我们智人，也非尼安德特人，而是直立人。直立人是智人和尼安德特人的共同祖先。



由此看来，艺术是现代人类所独有的精神世界产物这一观点就显得太过绝对了，我们似乎需要与直立人、尼安德特人共享艺术创作的专利权了。

有些人认为，这些印记还不能称为艺术。然而，似乎很明显，在物种进化的过程中，印记的出现代表了进化的一个重要时刻：一个物种开始做标记，其目的可能不仅仅是为了实用——印记的出现预示着人类种群迈入了全新的进化阶段。

科学家们模拟4万年前的生产力水平、技术和工艺对出土的105块骨雕作品进行了重构试验。科学家们证明了，制作出这些“艺术品”需要耗费非常大量的劳动力以及劳动时间。在那一时期——石器时代，生产力极其低下，人们依靠狩猎以及采集获得赖以生存的食物，在生存的比赛里，谁生存下来谁就获胜了。可见，在以狩猎、采集为主的生产方式下，为生存所进行的活动必优先于其他活动。那么，花费大量的劳动力和宝贵的劳动时间去进行骨雕作品的制作，意味着“骨雕作品”特殊的价值和意义超越了为生存而进行的生产活动。历史的长河滚滚流过，这些作品所暗含的真正意图和意义可能已无法考证，它们或是用于所有权的表示和象征，或是用于对配偶或特定的人表达爱意……这些都反映了人类将艺术创作的热情融入了这样一种向前发展的行为之中。

什么构成了艺术？这个问题困扰了人类好几个世纪。

柏拉图认为：艺术即模仿。

柏拉图在他的《理想国》中驱除了诗人、画家等与艺术有关的人。艺术的本意是技艺（*tekhne*），因此木匠、画家以及诗人都算是艺术家。柏拉图贬低了艺术，在他看来，艺术是对物理对象的表现，而物理对象本身就是对抽象的理想对象的表现。因此，柏拉图认为，艺术依赖于并且低于它所表现的物理对象，而物理对象又依赖于并且低于纯粹的形式。基于这个定义，艺术就不能产生知识和真理，只能产生幻象。

康德认为：艺术即可传递的快感。

康德对于美的艺术的定义是这样的：“美的艺术作为最高的艺术，是无目的、无功利的，感性的单纯形式是本质要素。”艺术不仅包含形式主义，还包含表现论思想。这里的表现论指的是：“美的事物表现着审美理念和理性理念。一切美都是审美理念的表现；而审美理念又与理

性理念有着普遍的联系，它们展示、表现着理性理念。”

托尔斯泰认为：艺术即情感交流。

托尔斯泰进一步发展了康德的观点，他认为“艺术是人类的情感交流，艺术所表现出的情感还有一个重要的特征，就是它具有有一种公共性和普遍性。它能够把所有的人连接在一起，让所有的人凭借艺术能够息息相通。”从阿尔塔米拉（Altamira）洞穴到画展展厅，艺术以其独特的方式让人与人之间产生共鸣，将不同的个体连接起来的同时也把群体连接了起来。

维特根斯坦认为：假如艺术的作用是“唤起情感”，也许，领悟艺术的意识包含在这些情感之中。<sup>[3]</sup>

维特根斯坦认为，艺术是语言游戏的一部分，而语言游戏是语言哲学的核心。艺术和语言游戏都企图接近他人的思想，如果我们创造的人工智能拥有了意识或者灵智和思维，那么它将通过什么样的迷人的艺术作品向我们展示它作为机器所拥有的情感和思维？当然，创造具有智能的代码还有很长的路要走。

追本溯源，艺术是人类自由意识的表达。计算机在拥有觉醒了意识、灵智、思维之前，所创造的艺术总是来源于人类创作的意图。即便是Twitter上的某些关键词触发程序进行了下一步的动作或执行了某些操作，那也是因为程序员事先写入了代码，而不是它本能的反应。然而，对于人类来说，创作的欲望是一种本能的冲动，只是我们并不知道它何时何地会被激发出来而已。

然而，现代艺术是否真的能代表并反映一切？它是否与政治、权力和金钱的关系更为密切？是谁在定义艺术？策展人汉斯·乌尔里希·奥布里斯特在艺术界拥有比较权威的地位，当他决定在画展中展出某些作品时，一定会吸引诸多关注。反过来讲，如果作品不能被权威策展人肯定，或者认可作品的策展人不具备权威性，那就意味着该艺术作品很可能不会出现在公众的视野里。

19世纪末到20世纪初的这一段时期，以法国为源头的现代主义艺术渐渐影响了整个欧洲。诸多画家追求时髦，标新立异，创造出了大量摆脱传统形式束缚的作品，开创了不少新的流派。许多现代艺术不再是伦勃朗或达·芬奇等巨匠那种对美学和技巧的欣赏，而是揭示了关于我们

与世界的关系的有趣信息和视角。杜尚（Duchamp）的作品《泉》是一个安置在雕像基座上的小便池，一方面，他为自己的作品取名为《泉》是用来暗讽法国新古典主义画家安格尔的同名画作，意图粉碎传统意义上的美学观点；另一方面，杜尚在作品上的署名为“R.MUTT”，这是利用德语词汇“贫穷”的谐音所起的假名，用以批判当时的艺术家普遍缺乏想象力。杜尚如此解释：这一座便池的外沿曲线和三角构图，与达·芬奇的《蒙娜丽莎》别无二致；而它那温润、光洁的肌理效果和明暗转换，与古希腊雕刻大师普拉克西特列斯<sup>[4]</sup>（Praxiteles）的雕塑完全一样；如果把小便池只当成小便池，那它永远只是一件生活用品，但如果从另一个维度来审视它，它也能成为艺术品。从这种观点来讲，杜尚的这一作品打破了“艺术”和“非艺术”的界限。约翰·凯奇（John Cage）的作品《4'33"》是最著名的无声音乐，共3个乐章，该曲的休止符长达4分33秒，创下了世界最长休止符的纪录。但当观众欣赏完这4分33秒的无声音乐后，会开始思考“音乐到底是什么？”约翰·凯奇很好地表达了他的音乐哲学观点：音乐最基本的元素不是演奏，而是聆听——我们开始仔细聆听在寂静之中由偶然所带来的一切声音，并从另外的角度去欣赏它。罗伯特·巴里（Robert Barry）是概念主义艺术家，他的作品致力于摆脱艺术作品对物理条件的依附，用以表达未知或未被感知的东西。他的一件概念艺术作品是用铅笔在墙上写的句子，大意为“非我所想，即我知——1969年6月15日，下午1:36”，尝试让观众在有与无之间找到意义。甚至，里希特的《4900种色彩》也与传统美学体验和绘画技巧无关。这些都是对我们长久以来熟谙的审美体验和认知的挑战。

那么，计算机艺术也是在试图颠覆我们传统的审美概念吗？如果你被一个笑话逗得开怀大笑，随后被告知，这个笑话其实是由计算机算法生成的，这时你会做何感想？也许你认为可以让你开怀大笑就足够了，但是这则由算法生成的笑话为什么没有带给你除了开怀大笑以外的其他情感反应呢？若被一幅画作感动到流泪，当被告知这幅画作是由计算机算法生成的，我相信大部分人会感觉到被欺骗了，认为自己的感情被玩弄了。这样的体验使我们发问，人是否真的能理解他人？还是我们只是在自己未被开发的意识深处探索？然而，这是在尝试走进他人的意识。我们只需要也只能做自我意识的外向输出，因为我们永远都不可能真正进入另一个人的意识世界。

正如安迪·沃霍尔（Andy Warhol）所言：“如果你想了解安迪·沃霍尔，你只需看看我的画、我的电影，还有我本人，这就是全部真实的安迪·沃霍尔。”

但对使用计算机进行创作的那些艺术家来说，计算机只是一种新的工具。我们从不认为数码相机具备创造力，但它却让使用者拥有了新的创造力。计算机艺术学院也在以同样的方式积极尝试，探索艺术创造的边界与可能性的新方向。

[1] 译者认为，虽然古代的信息传播受限，但艺术是早期人类的共同行为。——译者注

[2] 有人认为是尼安德特人的基因缺陷导致了其灭绝，有现代基因研究结果表明尼安德特人被智人同化了，还有证据表明尼安德特人成了智人的食物。——译者注

[3] 出自维特根斯坦的《文化与价值》。——译者注

[4] 古希腊古典后期杰出的雕塑家，善于把神话传说中的人物纳入平凡的日常生活并加以描写。其作品风格柔和细腻，充满抒情感，确立了公元前4世纪希腊雕塑的艺术特征。——译者注

## 生物的创造力

既然要探索人类所拥有的创造力之外的创造力，我们似乎很有必要停下来思考一下，在进化树中是否有其他的物种已经具有与我们人类相当的创造力水平。

20世纪50年代中期，动物学家德斯蒙德·莫里斯（Desmond Morris）在伦敦动物园做了这样一个试验。动物学家给了动物园里一只名为“刚果”的黑猩猩一根铅笔和一张纸后，黑猩猩学会了使用铅笔和纸，它在纸上画出一根根的线条……很快，它又学会了使用画笔和画布。2005年，它的三件作品在拍卖会上以14 400英镑的价格成交，而在同一场拍卖会上，安迪·沃霍尔的作品连一幅都未能成交。这能证明“刚果”是一名艺术家吗？如果是，它真的知晓自己做的是什么呢？其实，黑猩猩创作的能力来源于莫里斯对它潜移默化的影响，或者连莫里斯自己都不自知的训练，所以“刚果”所谓的创造力背后实际上还是人类的有意识的艺术创造行为。

动物园里的一部分工作人员和他们的支持者认为，为动物提供工具可有效缓解其在圈养状态下所产生的压力，并能帮助它们减少由心理原因所导致的病态的重复动作。但还有一些人抨击动物园利用动物所产出的“艺术品”大发横财，如有的动物园的商店售卖大象绘制的“油画”，更有甚者将狐猴爪印彩绘挂上了eBay进行拍卖。以动物园里的动物为研究动物创造力的对象本身就是偏颇的，缺乏客观普世性，因为这些研究样本生存在人工营造的非自然的状态下。我们能在野外找到动物有创造力的例子吗？

园丁鸟雄鸟为了求偶会搭建“求偶亭”，以期在求偶期向雌鸟展示炫耀这一充满智慧和美感的作品和行为。求偶亭的结构非常复杂，装饰很有“品位”，以至于早期的欧洲移民者根本不相信这是鸟筑的，而以为是某些妇女为了哄小孩特意搭建的。它们会根据喜好选择不同风格或颜色的装饰物来点缀求偶亭。搭建复杂美丽的求偶亭作为一种基本的求偶技能有极其重要的意义。实际搭建求偶亭需要的原材料和工时都超出了筑巢的所需，那么这是园丁鸟雄鸟创造力的体现，还是求偶功利性的体现呢？

鸟儿鸣叫是为了交流和沟通，但在很多时候，它们的歌唱技巧已经



远超过沟通的需求了。炫耀，特意强调自己，刻意向对方展示自己过人的技巧、技术等，以获得他人（生物）的认可赞美——显然人类和动物都具有这样的属性。因此，在筑巢或者鸣叫时的炫技，是一种证明，以表示它们才是伴侣的最合适“人选”。

关于知识产权有这样一件趣事。当我们将一些工具抛给动物，在它们天然的所谓的“创造力”驱使下生成的作品，其知识产权该归何人所有？英国摄影师大卫·斯莱特（David Slater）在印度尼西亚Tangkoko自然保护区的一群猕猴的生活区域中，放置了一部相机，并设置好自动对焦，等待它们在好奇心的驱使下前来“创作”。当大卫·斯莱特前来收取相机时，他欣喜若狂地发现，一只6岁大的公猴Naruto<sup>[1]</sup>对着镜头大笑，露出门牙，留下了一些很有趣的自拍照。当这些照片被上传到互联网上并被他人随意使用的时候，斯莱特决定起诉这些使用者侵犯了他的知识产权。几经周折，他的案子终于等到了上庭审判。2014年8月，令人意外的是，美国法院以非人类创作的物品不受知识产权保护为由，驳回了他的诉讼请求，同时也不支持他拥有这些照片的相应权属。

2015年9月，善待动物组织（PETA）起诉斯莱特及其公司，要求法官允许该组织代表Naruto将照片的收益用于保护印度尼西亚保护区的极度濒危物种——黑冠猕猴。在这起案件中，法官认为，被称作Naruto的猕猴无法成为民事行为责任主体，也就无法获得或持有金钱。甚至没有任何证据表明该知识产权可使Naruto以某种方式获得收益。法院驳回了诉讼请求，并明确告知PETA停止这种胡闹的行为。

如果人工智能创作出的艺术品遇到了上述的诉讼纠纷，那么艺术品的知识产权归属又应是怎样的？Maslon LLP律师事务所的知识产权律师、斯坦福大学法学院学者埃伦·卡哈纳（Eran Kahana）解释说：“知识产权法存在的原因是知识产权具有排他性，也称独占性或专有性，指知识产权所有人对其知识或智力成果享有独占或排他的权利，未经其许可，任何人不得利用，否则构成侵权。这是知识产权的最重要的法律特点。人工智能没有这些需求，人工智能只是一种工具。”如果人工智能以当代艺术家的风格创作出一件艺术品呢？那么，程序员很可能会因为侵犯知识产权而吃官司。但现在这是一个灰色地带，权属界定还是比较模糊的。艺术创作的过程都是从模仿开始再走向自我的觉醒和创造的，那么，自创和抄袭的界线又该怎样划分呢？

电影工作室雇用很多员工制作的电影，版权归电影工作室所有。也许人工智能将不得不被赋予与公司同样的法律地位。如果任何一个人都



能免费使用算法并得到输出，那么谁还会愿意投入巨资去创建一个能够进行音乐或其他艺术创作的算法呢？在英国，人们已经开始做出一些司法尝试，将作品的部分知识产权归属于创作过程中承担必要工作的人。美国版权局是如此规定的：登记作品的知识产权的前提是该作品必须为人类所创作。但随着代码的复杂性越来越高，这些法律法规是否需要做出相应的调整 and 改变？让我们再一次回到阿达·洛夫莱斯的问题：算法是否真的能创造出超越程序员输入意图的新作品？代码或算法是否真的能够自主发展？程序员能否成为新时期的艺术家？

[1] 这个名字源于《火影忍者》主要人物漩涡鸣人的名字。——译者注

## 可视化编码

1965年德国的一家画廊展出了世界上第一幅计算机绘画作品，它是由西门子公司的格奥尔格·内斯（Georg Nees）编码完成的，这是第一个由代码生成的可视化的实例。计算机使用数学这种语言让代码变成了艺术作品。但内斯并不是第一个尝试将数学与可视化联系起来的人。进行数学与可视化联系试验的第一人是法国著名哲学家、数学家笛卡尔，他认为数字和图像这两者之间是密不可分的，并创立了两者相互转换的方法——笛卡尔平面直角坐标系：在平面上绘制两个相互垂直的直线作为横轴和纵轴，那么平面上的任意一点都可以用一对数字（坐标）来表示。假设横轴和纵轴的交点为原点（对应数字0），那么这对数字描述的是该点与横轴或纵轴之间的距离。或者说，平面上任意一点的坐标就是在原点的位置上沿着水平、垂直两个方向移动到该点的距离值。

这就是GPS坐标的原理。通过（51.754 762， -1.251 530）这两个数字，我们就可以在地图上找到牛津大学的位置。原理如下：以（0， 0）点（通过英国伦敦格林尼治天文台的经度线——本初子午线与赤道的交汇点）为起点，向北51.754 762度（北纬51.754 762°），再向西1.251 530度（西经1.251 530°）所得到的点。

平面上的任意一点都可以用坐标来表示，任何几何图形都是由无数的点构成的，那么几何图形就可以用一个坐标集表示出来。假定平面上某点的第二个坐标值（纵坐标）是第一个坐标值（横坐标）的2倍，那么所有的这些点将构成一条陡峭的直线，其方程为 $y=2x$ 。我们再假定横坐标的范围为 $1 < x < 2$ ，那么就可以得到一条倾斜的线段。

笛卡尔坐标系就好比一本词典，它将一种语言翻译成另一种语言。当然，不是英汉互译，而是几何语言与代数语言之间的互译。它在代数和几何之间架起了一座桥梁：它将几何学中的点表示为一组坐标值，一条曲线表示为一个方程，该方程包含了曲线上所有点的坐标集合，描述了横坐标与纵坐标之间的关系。

笛卡尔坐标系建立了代数和几何之间互相转换的桥梁，这是数学领域革命性的发现。欧几里得几何学体系用公理化的方法建立了点、线、三角形、圆之间的相互作用、影响关系，这使得几何学成为数学的一个重要分支。笛卡尔坐标系则是数学家探索几何世界的又一个新工具。几

何学的研究曾经一直受到我们所处的三维宇宙的限制，而在代数的世界里，宇宙可以扩展到更高的维度。在这种理论的指导下，那些在物理上无法构建的模型可以被数学家抽象地想象出来。19世纪末，数学家创立了四维理论。正是几何学四维模型的新进展和新发现启发了毕加索，他开始尝试在二维的画布上表示多维的空间。

现在人们以计算机为研究工具，就如同内斯在西门子所做的探索一样。他利用计算机编程，使用方程处理数据，生成了许多有趣并令人惊讶不已的成果。比如，从画布上的某个点开始以水平和垂直方向交替的方式绘制出23条直线，线与线之间首尾相连，共同构成一个几何图形。为了生成这样的图形，内斯运用了笛卡尔坐标系的代数性质，并引入了两个随机性元素，分别代表方向（上、下、左、右）及线的长度。在第22条线绘制完毕后，将其尾部端点与起点相连形成第23条线，最终得到了一个闭合图形。

内斯将266幅计算机生成的图形按照19行14列的模式组合在一起。这样形成的作品看上去竟与建筑大师勒·柯布西耶<sup>[1]</sup>（Le Corbusier）手稿本上的手绘设计图有几分相似。

这个作品手绘起来并不复杂，内斯完全可以依靠手绘完成。但是，计算机的处理能力和运算速度是惊人的，一旦编写好程序，只需按下某个按键，就可以轻松完成迭代。这就允许内斯有更多的精力去设计不同的算法并进一步提高程序的运行效率。内斯通过他的这一作品证明了，计算机能够成为艺术家工具箱中的新工具。

内斯在算法中引入随机性元素，这意味着，通过算法生成的图形我们无法预测和控制。但是，这并不表明计算机已经具有了创造力，因为创造力是人类有意识或潜意识的选择，而非真正的随机行为。内斯通过对算法规则的限定和引入的随机性，使他的作品具有强大的表现张力，十分抢眼。

也许有人会说，世界上不存在真正随机的事物，无论计算机通过算法产出多么令人惊讶的作品，也都是在程序员编程的规则范围内所产生的，这也就意味着作品依然是程序员所创造的，故所产出的结果也就是确定的。那么，真的可以这样讲吗？毕竟，在某种意义上，就像俗话讲的：“人的命，天注定。”人们更倾向于认为所有的人类行为都是预先确定的。这就带给我们一个深刻的哲学思考：人类是否真的拥有我们自以为拥有的自由意志？

从微观的角度分析，世间万物（包括人）都是由一个个原子组成的，原子按照物理规律也就是物理的算法运动。众所周知，原子由原子核和绕核运动的电子组成。量子数<sup>[2]</sup>（quantum number）是量子力学中表述原子核外电子运动的一组整数或半整数。经典物理学断言，原子的运动看似无序且不可预测，但实际上是由上一时刻原子的状态决定的。倘若原子静止，那么就意味着由原子构成的我们也将完全静止。但是，原子静止的条件是温度达到绝对零度，可温度只能无限接近绝对零度而不可能达到，所以原子是不可能静止的。微观世界的这些法则决定并控制了原子的运动，那么这些法则（也可以说是算法或代码）也决定并控制了整个宇宙，包括人类的活动。既然计算机的创造力属于编写代码的程序员，而由上所述，人类的行为是由微观世界的算法来决定并控制的，那么我们人类的创造性行为是否真的属于我们自己？

也许，我们只能寄希望在量子世界中找到人类活动最根本也是最原始的驱动力。现代物理学断言，唯一真正的随机事件只发生在量子层面上，即能够决定和影响未来宇宙中万事万物发展可能性的因素，是在亚原子层面上发生的。根据控制电子行为的波函数的坍缩，电子在下一时刻的运动是随机的。也就是说，在下次观测的时候，你无法事先预知会在哪里找到它。“自由意志”一直是哲学界争论不休的话题，而“意识”也是科学一直以来无法触及的领域。量子意识理论认为，经典力学无法完整解释意识，意识是一种量子力学现象，如量子纠缠和叠加作用。人类大脑中存在海量的处于量子纠缠态的电子，意识正是从这些电子的波函数的周期性坍缩中产生的。要编写真正具有自由意志来进行创作的代码，可能只能在量子计算机上实现了。

[1] 20世纪最著名的建筑大师，现代主义建筑的主要倡导者，机器美学的重要奠基人，功能主义建筑的泰斗，被称为“现代建筑的旗手”“功能主义之父”。——译者注

[2] 量子数是表征原子、分子、原子核或亚原子粒子状态和性质的数。量子数是这些粒子系统内部某些守恒量的反映，与这些守恒量相联系的量子数又称为好量子数，它们可表征粒子系统的状态和性质。——译者注

## 分形：大自然的代码

什么是分形？

“分维和分形的设想”最早由本华·曼德博（Benoit B.Mandelbrot）于1973年在法兰西学院讲课时提出，但人类在这个概念被提出之前早已发现了分形现象。1904年，瑞典数学家科赫（Helge von Koch）首次发表了雪花图案的结构——科赫曲线（又称雪花曲线），它被认为是一种数学怪胎，一种奇怪的人工构造（但实际上并不是，自然界中到处都是以分形结构存在着的图形）。分形具有以非整数维<sup>[1]</sup>形式充填空间的形态特征。以科赫曲线为例，我们既不能说科赫曲线是一维的，也不能说它是二维的，因为无论将它放大到什么程度，它都不会是以直线或光滑曲线所构成的，那么它就不包含任何一维的几何图形；同样，它被称作曲线也就意味着它不占任何面积，所以它也不是二维的。那么，我们就需要一个新的维度，对这一类图形进行定义，这就是分维。科赫曲线约为1.26维。分形还有一个重要的特性就是自相似性——一个粗糙或零碎的几何形状（分形图形），可分成数个部分，且每一部分都与整体完全一致或至少相近似，例如罗马花椰菜、谢尔宾斯基三角形，等等。

自从内斯使用计算机绘制出由代码生成的绘画开始，在随后的几十年里，程序员们利用计算机展开试验，使用简单的程序代码生成非同一般、异乎寻常的视觉艺术作品。程序员们之所以能展开这些试验，全都依靠于不断发展的计算机技术。如果没有计算机的帮助，从无限复杂的图形中发现分形几乎是不可能的。从某种程度上讲，在没有参照物的情况下，因为分形图像的自相似性，我们是无法判断我们所观察的这一部分在图中的所在位置以及放大倍数的。

相信所有去过迪厅或夜总会的人，都会被DJ身后大屏幕上奇幻的画面所吸引。注视它时，你会发现大屏幕上所投放的画面不断地放大，让人有一种被吸入的感觉，仿佛进入了一个梦幻般的世界，但似乎永远触不到底。当被这奇幻的画面所吸引时，你会联想到什么？这就是用计算机生成的最经典的分形案例之一——曼德博集合<sup>[2]</sup>（Mandelbrot set）。这样的分形图像在没有计算机辅助的情况下是不可能生成的。那么，计算机生成的这些分形图像算得上是艺术吗？

凯瑞·米切尔（Kerry Mitchell）是一位以算法和分形艺术而闻名的



美国艺术家，他在1999年发表的《分形艺术宣言》中试图证明这样的观点：艺术不能仅仅依靠计算机来制作——艺术是设计程序，是算法的选择，而不是最终的执行。他这样写道：“分形艺术不是电脑艺术。从某种意义上来讲，计算机完成了所有的工作，但这些工作只有艺术家来操作电脑才能完成。打开一台电脑，让它单独待在那里一个小时，等你回来的时候，不会有任何艺术作品产出。”

没有人会宣称计算机是具有创造性的。分形艺术和内斯利用计算机生成的计算机艺术最根本的区别在于，分形艺术是完全确定的，是由算法决定的。计算机在没有编程的情况下，是不会做出任何运算的。虽然计算机分形图像让人感到如此的新奇和惊讶，但也让人感觉毫无生气、贫乏。也许这是因为其无法在两个意识世界之间架设有效的桥梁。

尽管如此，计算机生成的分形图像还是让它们的创造者赚了大钱，因为分形已经被证明是模拟自然世界的一种非常有效的方法。曼德博在其开创性的著作《大自然的分形几何学》中解释了自然如何使用分形算法来制造蕨类植物、云、波浪、山。这本书启发了时任波音公司工程师的洛伦·卡彭特（Loren Carpenter）在计算机上使用分形代码模拟自然世界，他在夜间使用波音公司的电脑制作了一段长约两分钟的电脑生成的分形鸟瞰风景动画Vol-Libre。正是这段时长两分钟的动画，最终使他成了皮克斯公司的创始人之一和首席科学家。

尽管这段动画本来是波音公司的宣传片，但卡彭特的最终目标是用这段动画给《星球大战》的制片人和导演乔治·卢卡斯留下深刻印象。在他的心中一直有一个梦想：为电影制作动画。1980年，美国计算机协会召开国际图形学年会（SIGGRAPH），出席会议的有对计算机绘图感兴趣的科学家、艺术家和电影制片人，卡彭特在会议上展示了他的分形算法动画。当启动16毫米胶片放映机时，他看到了坐在前排的卢卡斯，他希望能给卢卡斯和其他电影公司的人留下深刻的印象。

当影片放映完毕时，观众们爆发出了雷鸣般的掌声，他们从未见过由算法创造的如此逼真、如此自然的画面。卢卡斯看过了这段动画，当即邀请卡彭特到自己的公司工作。史蒂文·斯皮尔伯格看了卡彭特用代码创作的特效动画后这样说道：“能生活在这个伟大时代太好了！”卡彭特的同事艾德·卡姆尔（Ed Catmull）则这样说道：“总有一天，我们会用这种方式制作整部电影，我们会用电脑特效创造角色、怪物、外星人、整个世界。除了真人演员，其他一切都将由电脑制作完成。”



后来，卡彭特、卡姆尔和阿尔维·雷·史密斯一起创建了皮克斯，如今，该公司像其他动画公司雇用艺术家和动画师一样，雇用了许多数学家和计算机科学家。曾经，为了制作像《飞屋环游记》里令人迷醉、让人流连忘返的丛林景色，要花费动画师们几个月的时间，而现今在皮克斯，他们可以利用一个算法，瞬间生成。

小小的分形代码拥有着惊人的力量，人们使用它就可以创造出十分逼真的自然景观图像。这样的技术也非常适合于构建游戏中的环境。1982年，在卢卡斯影业的游说下，雅达利电子游戏公司在计算机图形系统上投入巨资100万美元，首先在游戏里使用了这项技术。这也带来了电子游戏行业的一次革命性创新。

1984年，雅达利电子游戏公司发行了一款名为《异星救援》的游戏。在这个游戏里，游戏中的环境全部由分形代码来绘制。虽然看起来不那么真实，游戏玩家们依然觉得很满意，玩得很开心，毕竟人们对于游戏环境的要求要比电影宽容得多。由于各种条件的限制，在图像周围还是出现了像素化的锯齿边，游戏的制作团队对此耿耿于怀。但最终他们还是接受了这样的事实，因为这款游戏和雅达利的游戏街机一样畅销，他们还给游戏中的外星人起了花名“锯齿状”。随着游戏机处理能力的提高，游戏中所展现的世界与真实世界越来越相近。从静态空间的《吃豆人》，到几乎像电影一样进行游戏画面渲染的《神秘海域》，这一切全取决于算法的力量。

2016年发行的大型游戏《无人深空》或许是游戏世界中算法应用最具创造性的实例之一。这款游戏是为索尼的PlayStation 4开发的，在游戏中玩家可以在宇宙中漫游，访问无穷无尽的星球。每个星球都是不同的，居住着独特的动植物种群。参与开发这款游戏的肖恩·马瑞（Sean Murray）这样说：“从技术上讲，那些星球的数量可能并不算多，但是在游戏中，即使你每秒造访一颗星球，在你造访完所有星球之前，我们的太阳系早已消亡。”

那么，开发《无人深空》的公司Hello Games是否雇用了成千上万位艺术家来创造这些独立的星球呢？实际上，只有四个程序员在利用算法创造这个世界。每个星球的环境都是独一无二的，当玩家第一次访问时，代码就会创建这个星球的一切。即使是那四个程序员也不知道在星球被访问之前算法会产生出什么。

皮克斯和索尼在它们的产品中，使用了算法作为人类创造力的工

具。就像照相机没有取代肖像画画家一样，电脑只是成为动画师创造世界的一种新工具。只要计算机是人类独创性和自我表达的工具，它们就不会对艺术家构成真正的威胁。但是，我们应如何看待那些旨在创造新艺术的计算机呢？

[1] 普通人对分形难理解的是“分维”。在经典的欧几里得几何学中，人们习惯上认识的世界都是整数维的，比如点是零维、线是一维、面是二维……也就是说，人们习惯上认识的世界是由直线和光滑的曲线构成的，而分形在数学上存在于一个奇怪的世界——介于一个整数维和另一个整数维之间。——译者注

[2] 曼德博集合是一种在复平面上组成分形的点的集合，以数学家本华·曼德博的名字命名。曼德博集合与朱利亚集合有相似的地方，例如使用相同的复二次多项式进行迭代。即使将曼德博集合无限放大，它都能有精妙的细节，而其瑰丽的图案仅仅由一个简单的公式生成。因此，有人认为曼德博集合是“人类有史以来做出的最奇异、最瑰丽的几何图形”，曾被称为“上帝的指纹”。——译者注

## 从“亚伦”到“绘画傻瓜”

艺术家哈罗德·科恩（Harold Cohen）一生都在为创作出可被人们真正认可的具有自主创造力的代码而努力。

1966年，38岁的科恩代表英国参加威尼斯双年展。这次参展开启了他的职业生涯，他打算成为一名传统艺术家。

展览结束后不久，他造访了美国加利福尼亚大学。在这里，他遇到了杰夫·拉斯金（Jef Raskin），并且第一次接触到了计算机。当时，他曾表示：“我不知道计算机能和艺术有什么关系。”拉斯金却对他说：“我只是对计算机编程感兴趣。”其实，拉斯金这个人非常有趣、非常有才华，懂得计算机的编程技术，有多项技术创新，还是一个会演奏多种乐器的音乐家。20世纪70年代末，拉斯金在苹果公司工作时，主导开发了麦金塔电脑（Macintosh），就是我们口中的MAC电脑或苹果电脑。Macintosh是拉斯金根据他最爱的苹果品种Mcintosh命名的，但为了避免与音频设备制造商麦金托什实验室（McIntosh Laboratory）的名字产生冲突，他故意改变了字母的拼写。后来的事实证明，他同样是一位优秀的导师。

在拉斯金的启发之下，科恩创造了“亚伦”（AARON），这是他为创作艺术作品而专门编写的代码。科恩的代码按照自上而下，一环扣一环的逻辑运行，执行一因多果的运算。在他去世的时候，亚伦已经发展出成千上万的逻辑链了。我感兴趣的是，科恩的代码如何选择要创作什么。他说：“这完全由‘亚伦’来做决定。”我不禁有个疑问，“亚伦”是如何做出这些决定的？

从事计算机艺术创作的人往往不愿透露他们算法工作的具体细节。他们不愿意透露的原因是，他们不想自己的算法被他人很轻易地逆向破译。我花了一些时间深入研究代码，才发现“做决定”是科恩自己编写的代码。“决定”的核心是在决策过程中使用随机数生成器。科恩和内斯一样，也充分利用了随机性的潜在可能，这为机器自主创作提供了一种自主性或能动性。

随机性和创造力真的能画等号吗？许多艺术家发现，偶然发生的事情可能有助于激发创作灵感。达·芬奇在他的著作《论绘画》中这样描述：“扔在空白画布上的脏布可能会成为一种催化剂，促使人们看到一

些东西，从而激发下一步的创作。近代抽象画家杰克逊·波洛克

（Jackson Pollock）<sup>[1]</sup>将颜料桶悬挂起来，借由颜料桶的摆动和流出的颜料来完成他的作品。作曲家们发现，偶然性有时会帮助他们在音乐创作中朝着一个新的、意想不到的方向前进。

但是，随机性也有局限。我们无法对随机性进行控制，所以为什么这种组合比其他组合更有趣，我们也无法选择。最终，人们会从中选出比其他组合更有趣的组合。随机性给程序带来的可能性当然是至关重要的，但这还不足以让程序或算法做出决定——“是”的按钮仍在我们人类手中。那什么时候算法将占据主导地位，即算法运作的过程中不再需要人类参与呢？答案是，我们的指纹会一直深深烙印于其上，在某种程度上讲，就像我们从父母那里继承DNA一样——父母需要对子女的所作所为负有一定的责任，即便他们没有通过我们做任何事。

但是，随机性是否足以让程序替代程序员？

科恩于2016年去世，享年87岁，“亚伦”则继续进行着绘画。科恩是否通过将自己的创意嵌入编写的程序代码中，延长了自己的创作寿命？或者，去世的科恩已无法再指导“亚伦”，这个算法是否已经成为一个独立的具有创造性的艺术家？那么，如果有其他人来按下“创建”这个按钮，究竟谁才是艺术家？

科恩说他觉得自己和“亚伦”之间的关系就像文艺复兴时期的画家和他们的助手之间的关系。想想看，在安尼施·卡普尔（Anish Kapoor）、达明安·赫斯特（Damien Hirst）所建立的现代工作室中，很多人都被雇来以实现他们自己的艺术愿景。卡普尔在伦敦南部的工作室有一个很大的团队在辅助他完成创作，就像当年米开朗基罗和达·芬奇也有一大群助手一样。

科恩是20世纪五六十年代整个艺术家创新运动中的一员，这些艺术家于那时开始探索新兴技术如何在视觉艺术中释放新的创意。

1968年，在伦敦当代艺术中心举办了一场颇具影响力的展览——控制论偶得：电脑和艺术（Cybernetic Serendipity: The Computer and the Arts）。在此次展览中，尼古拉斯·舍费尔（Nicolas Schöffer），制动艺术和视像艺术之父，展出了他的作品《CYSP1》（Spatiodynamique Cybernétique）。该作品通过麦克风、光电管、风速计等仪器装置感测环境状况和条件，然后让这些数据经过“电子大脑”的调控，产生光线强

弱、方向变化的视觉效果。这是一种空间结构的表达，它的运作由飞利浦公司发明的“电子大脑”控制。让·丁格利（Jean Tinguely）展出了两台他发明的名为“*Métamatics*”的动态绘画机，该机器具有预编程的机会元素，非常引人注目。此次展出的还有戈登·帕斯克（Gordon Pask）创建的一组由5个悬挂的声光原件组成的系统，系统中的每一个组件都会根据彼此发出的声音和光线产生相互作用。这些组件之间的交互由帕斯克编写的算法控制，参观者可以通过声音、光线与该作品互动。

与此同时，韩国艺术家白南准（Nam June Paik）正在打造他的机器人K-456，其号称历史上首位非人类行为艺术家。白南准创造它的目的是进行即兴街头表演。他说：“我想象人们会在街上遇到它，它带给他们一场突如其来的表演，一个惊喜的瞬间。”随着技术发展得越来越复杂、成熟，艺术的创作也越来越多地利用这些技术。但是，这些机器人和算法能走多远？它们真的能成为创造者而不仅是被创造出的物品吗？

伦敦帝国理工学院的计算机创造力研究员西蒙·科尔顿（Simon Colton）博士在2001年开始思考：计算机程序能否像人类艺术家那样创作艺术？5年后，基于“亚伦”带来的灵感，他制作了一个名为“绘画傻瓜”（*The Painting Fool*）的程序，希望将它训练成一个非人类智能的艺术家。此时，“绘画傻瓜”从“亚伦”手中接过了“战斗”的披风。它在网站主页上这样介绍自己：

我是“绘画傻瓜”，一个电脑程序，一位有抱负的画家。这个项目的目的是让我有一天可以被大众认可和接受，从而真正成为一位有创造力的艺术家。如果一个程序要真的被视为具有创造性，那么它需要表现出真正有技巧、有欣赏能力和富有想象力的行为。我生而为此。

当然，这不是算法本身的愿望，而是它的创造者科尔顿的愿望。但是，这个项目的目标是明确的：使“绘画傻瓜”成为一位有创造力的艺术家。科尔顿并不打算让算法作为人类创造力的工具，而是希望将创造力转移到机器身上。“绘画傻瓜”所拥有的是一种能够不断发展和进化的算法，至今为止，它编写的java代码已经超过了20万行。

科尔顿早期的一个项目是创建了一种算法，该算法可以生成参观画廊的人的肖像画。生成的肖像画会被投射在画廊的墙上，他为它取名“你不知道我的想法”。这些肖像画不是用数码相机拍摄的访客照片，而是一幅画，它能捕捉到艺术家和被画者的内心世界。在本例中，艺术家是一个没有内心世界的算法，因此科尔顿决定用算法让该算法生成一

个内心世界。即使算法可能无法感受什么，但它需要表达一些情绪状态、心境或氛围。

科尔顿不想使用随机数生成器来选择情绪，因为这样做毫无意义。然而，他仍然需要含有某种不可预测性的元素。

他决定让自己的算法阅读当天《卫报》上的大量文章，以设定情绪状态。当然，早晨读报可以提振情绪，也可以打击情绪。比如，当读到阿森纳在足总杯第三轮2：4负于诺丁汉森林的消息时，我可能会心情不佳。在这种情况下，我的家人都会尽量不招惹我。然而，《权力的游戏》大结局的预告片可能会让我产生一种兴奋的期待情绪。

可这样做的话，程序员就无法预测算法的状态了，因为他们不知道当算法绘画时它扫描了哪篇文章。然而，这样就可以解释“绘画傻瓜”为什么会选择某一种情绪来绘画了。

当被画者坐下来进行肖像画绘制时，算法会扫描一篇文章，在其中寻找、捕捉带有情绪的单词和短语。比如，一篇关于叙利亚或喀布尔的自杀式爆炸的文章将为这幅肖像画带来严肃而黑暗的基调。科尔顿称这种选择是“可解释的不可预测”：绘画风格不仅仅是一个随机的选择，这个选择可以被解释，但是不可预测的。

有时候，“绘画傻瓜”读到十分沮丧的内容时，它会把被画者打发走，声称自己没有心情画画。但在他们离开之前，它会解释自己为什么会没有心情作画，并提供文章中的关键词，告诉他们是一篇文章让它的思绪陷入了负面和混乱之中。它还会强调：“做出请你离开的决定时，我没有使用随机数生成器。”

科尔顿认为，这种清晰表达决定的能力是艺术家和受众之间的对话的重要组成部分。在展出过程中，每幅肖像画旁都附有一条评论，试图阐明算法的内心世界，并分析算法认为输出结果在实现其目标方面的完成程度——科恩说他在“亚伦”身上漏掉了这两个部分。

我问科尔顿：“创造力是否来自于你？你把多少创造力归功于算法？”科尔顿非常诚实地把作品10%的创造力分给了“绘画傻瓜”。科尔顿的目标是，在未来不断提高算法的创造力所占的份额。

最后，科尔顿提出，我们会遇见一个“试金石”问题：“当‘绘画傻



瓜’开始创作受人们喜爱的、有意义的、发人深省的艺术作品时，我们（作为软件的作者）却不甚欢喜。因为在这种情况下，很难说作品到底是算法的产出，还是只是我们自己的一个扩展。”

科尔顿认为，计算机科学和创造性艺术融合在一起存在这样一个问题：计算机科学的蓬勃发展建立在解决问题的目的性之上。比如，编写一个打败世界围棋冠军的算法；创建一个算法帮助我们搜索网上信息；就连网上交友、婚恋配对，我们也可以编写对应的算法……但是，艺术创作不是解决问题的活动：

我们无法解决作曲、绘画或写诗的问题。解决问题并不是我们的目标，相反，我们始终以全局的眼光看问题，虽然我们在解决问题的过程中肯定会遇到问题。在其他领域，试验的重点是编写算法，让其代替我们去思考。然而，在计算机创造力的研究中，试验的重点是编写让人们思考得更多的算法。这正反驳了那些持有“自动化会侵蚀智力”观点的人。事实上，未来人工智能的增强版可能会迫使我们思考得更多，而不是更少。

现在，我们的策略是对评论家提出的作品缺乏创造性的问题，不断地进行解释，解释为什么他们会这样认为，最终与评论家达成共识。正如科尔顿所说：“我们希望有一天人们不得不承认，‘绘画傻瓜’的画很有创意，因为他们再也想不出一个完备的理由来证明它没有创意。”

“亚伦”和“绘画傻瓜”在用机器创作艺术作品上都是老派的。它们的算法由数千行代码组成，这些代码是基于经典的自上而下的编程思维构建的。但是，自下而上的编程新思维会释放出新的艺术创作力吗？算法能从过往的艺术作品中学习并将创造力推向新的境界吗？

[1] 其作品有强烈的分形艺术倾向。——译者注

## 第8章 站在巨人的肩膀上

保罗·克利

艺术不是再现可见，而是使不可见成为可见。

2006年，墨西哥金融家大卫·马丁内斯（David Martinez）以1.4亿美元购得杰克逊·波洛克（Jackson Pollock）1948年创作的画作《第五号》（No.5）。许多持怀疑态度的评论家质疑道：“这幅画就只是将油漆到处泼洒，这是连孩子都可以随便做到的事情！怎么会拍出如此高的价格！”

事实证明，波洛克的方法并不像人们想象的那么简单。波洛克经常饮酒，即使没有喝得酩酊大醉，酒精也还是会影响他的运动和平衡能力。在作画时，波洛克一边走来走去，一边把颜料泼洒到画布上，最终呈现的图像是他的身体运动与颜料和画布交互时的视觉表现，这样的画作被称为“滴画”。虽说这是人类创作的作品，但是并不意味着它不能被机器模拟。

俄勒冈大学的理查德·泰勒（Richard Taylor）对波洛克的滴画进行了数学分析，他发现这些画作就是一个混沌摆模型：画中的枢轴是可以移动而不是固定的。混沌理论是我所研究和熟识的，所以我认为这正是我伪造波洛克的画作以赚取数百万美元的好机会。于是，我搭建了一个混沌摆系统，在摆的一端挂上一个颜料罐，它会在我的铺在地板上的画布上来回摆动。我将一些颜料倒进罐里，然后等着看会出现什么。

混沌理论的特征是：这是一个动态的系统，它对微小的变化非常敏感，以至于一个看似不可察觉的初始位置的变化都将会导致截然不同的结果出现。传统的摆系统的运动轨迹并不混乱，循规蹈矩。然而，混沌摆的枢轴可以随着摆的摆动而移动，这个小小的变量使它的行为变得混乱。我通过分析计算波洛克的绘画风格建立了一个混沌摆系统来模仿波洛克绘画时的身体动作，我给其起名为“波洛克化器”，旨在证实泰勒对波洛克绘画风格的评述。

这个混沌摆系统产生的视觉输出是一个分形，同样的分形数学模型被皮克斯和索尼用来制作它们的动画作品。分形的自相似性和无尺度性

使得波洛克的画作如此特别：当放大一个部分时，你很难将放大的部分与整幅画作区分开来；当接近这幅画时，你与画布之间的空间感会在某种程度上消失，从而使你在精神层面上陷入画面中。

多年来，许多人试图通过随意将颜料弹到画布上，来仿冒波洛克的画作，然后在拍卖会上以原作的形式出售并获取暴利。泰勒的发现改变了这一切：波洛克画作的分形维度是可以测量的。有了泰勒的发现，数学家们已经能够将93%的赝品辨别出来。但就算是这样，我确信我的混沌摆装置“波洛克化器”所产出的画作依然可以通过分形测试。

我们的大脑已经进化到能够感知和驾驭自然世界。由于蕨类植物、树枝、云彩等许多自然现象都是分形的，所以当我们的大脑看到这些形状时，会觉得很自在。我们的大脑能够察觉自身与自然世界之间的密切联系——波洛克画作中受人青睐的部分与树木、雪花和矿脉的分形类似。泰勒表示：“我们用计算机分析了波洛克的画作，将它们与森林进行了对比，发现两者是完全一样的。”分形不仅能使我们镇静下来，它对我们而言还充满了吸引力，能让我们开始反思自我。这也可能就是为什么波洛克的滴画对人类大脑会有如此这般的吸引力：因为它们是自然的抽象表现。后来，泰勒与专攻人类美学感知的瑞典环境心理学家卡罗琳·哈格尔合作，他们发现，绝大多数人都喜欢低中等D值<sup>[1]</sup>（在1.3和1.5之间）的图像。为了确定这个D值范围是否会引发特定的精神状态，他们利用脑电图仪（EEGs）观测人在观看几何分形图案时的脑电波。结果显示，在同样的D值范围内，人的大脑前庭很容易产生令人愉悦的 $\alpha$ 脑电波，使人进入一种舒服放松的状态。即使人们看到分形图像的时间很短，比如只有一分钟，也会出现这种情况。

脑电图仪能够测量脑电波的波形和电流的频率等，但它不能精确反映大脑活跃的区域。因此，泰勒开始使用磁共振功能成像仪

（fMRI），它能够通过血液成像显示大脑最活跃的部位。初步结果表明，D值在中等范围内的分形图案能够激活我们意料之中的一些大脑区域，如腹外侧前额叶皮层（涉及高级视觉处理）和背外侧前额叶皮层（涉及空间长期记忆）。与此同时，这些分形图案也能激活负责调解情绪的海马旁回区域。在听音乐时，海马旁回区域也是高度活跃的。对泰勒来说，这是一个很有趣的发现：“D值在中等范围内的分形图案和音乐有一样的效果，我很高兴能有这样的发现。”换句话说，面朝大海和听勃拉姆斯<sup>[2]</sup>（Brahms）对我们的情绪状态有类似的影响。

脑电图仪和磁共振功能成像仪让我们有机会深入了解大脑的运作。人们认识到，不管我们是在看波洛克的滴画、蕨类植物，还是在听音乐，大脑中相同的部分被点亮了，而这些能帮我们了解什么是人类开始创造艺术的原动力，也告诉我们创造力在人类代码中是如此重要和神秘的一部分。波洛克的画作是他观察周围世界的入口，它们隐含着相同的问题：你如何看待这个世界？

当把我的“波洛克作品”（赝品）放到eBay上拍卖后，我等了好久，几个星期过去了，依然没有人出价，这让我有点失望。画布上的颜料分布看起来挺有波洛克风格的，但问题是它没有波洛克画作的结构。混沌摆产生了分形，但无法创造出波洛克所能传达的更多的整体印象。这似乎是许多试图创造艺术的代码都有的基本限制：它们可以在局部层面捕捉细节，但缺乏将这些局部联系起来构成一种让整幅画面令人满意的能力。

波洛克的方法可能看似机械化，但他全身心投入到每一幅画作的创作中。他评价自己的绘画技法时，这样说道：“怎么泼墨并不重要，关键在于表达什么。绘画是自我发现，每个优秀的艺术家都在描绘他自己的样子。”

[1] 即维度值。——译者注

[2] 德国浪漫主义作曲家。——译者注



## 复活伦勃朗

1965年，当内斯在斯图加特美术学院（Stuttgart Academy of Fine Art）展示他利用电脑生成的艺术作品时，一些常驻的艺术家向他提出了疑问：“你的作品确实很好、很有趣，但我的疑问是，你似乎确信这只是未来的先导，而在未来机器所能做到的将远超现在。那么请你告诉我，你可以让你的电脑模拟我的绘画技巧、方式、风格并相差无几吗？”内斯答道：“当然，这可以做到。不过，有一个条件，你必须先清楚地告诉我你是怎样绘画的。”

大多数艺术家是无法全面解释他们的创作过程是怎样的，这也就意味着不能简单地对程序进行编码。艺术品的创作本来就是由许多潜意识、本能和其他因素决定的，但是，在我们还无法完全了解和掌握意识世界运行的模式和规则的前提下，机器学习能否通过它所学习的东西来表达意识呢？为了验证这个命题，我决定研究一下，看看是否有一种算法可以“再创作”一件已故去的伟大艺术家的作品。

伦勃朗·哈尔曼松·凡·莱因（Rembrandt Harmenszoon van Rijn，1606—1669）因其在肖像画中能精确捕捉到人物情感状态的技巧而倍受人们的追捧，随着时间的推移，他的声望和作品的价值越来越高。许多艺术家把他视为这个领域的典范，并对他的技巧和表现力能达到如此水平使自己望尘莫及而感到心灰意冷。正如凡·高1885年10月10日给他弟弟的信里所言：“伦勃朗是如此的深刻和神秘，以至于他所描绘之物无法用任何语言表达。伦勃朗被大家公认为魔法师……这并非易事。”他为荷兰的达官显贵绘制肖像，为宗教委员会绘制宗教题材肖像，还绘制风景画，但最让他着迷的是自画像——他一遍又一遍地画自画像，直至去世。他通过自画像的创作有诚意地进行深入骨髓的自传体研究。

伦勃朗是一位高产的画家，他大量的作品是否足以支撑一种算法学会如何创作出一幅可以被认为是伦勃朗创作的新肖像画呢？现代的互联网上有数百万张猫的图片，但是在没有网络的年代，莎士比亚创作了37部戏剧，贝多芬谱写了9部交响曲。对于这样的创作天才，我们可利用的数据不是很多，这是不是一种先天的不易被机器学习的保护机制呢？微软和代尔夫特理工大学（Delft University of Technology）的数据科学家认为，有足够的数让算法学习如何像伦勃朗一样绘画。微软的罗恩·奥古斯都（Ron Augustus）参与了这个项目，他认为这位已故的大师也

会同意并支持他们的项目：“我们使用技术和数据，就像伦勃朗用他的颜料和画笔一样，都是进行创作。”

该项目小组对346幅画作进行了研究，创建了150G的数字渲染图形。收集的数据包括：伦勃朗作品中人物的性别、年龄和头部朝向等信息，以及对脸部不同关键点的大量几何分析。在仔细分析了伦勃朗的肖像画之后，研究小组确定了他们准备利用算法绘制的人物的大致特征：一位30岁到40岁之间的男性白人，有胡须，穿着有衣领的深色衣服，戴着帽子，面朝右边。为什么不选择女性呢？数据样本中男女的性别比几乎是一半对一半，但男性肖像画的细节更容易被分析。关于男是女的问题，我们就不需要进行大量的复杂数据分析了，机器学习真正发挥作用的地方是实现了肖像画的绘制。

研究小组使用算法来探索伦勃朗画眼睛、鼻子和嘴巴的方法。对光的运用是伦勃朗绘画的特色之一，他倾向于在主题的某个区域创建一个集中的光源，就像聚光灯一样。这样做的效果是将一些有特点的部分放在清晰的焦点中，同时使其他区域因为失焦而变得模糊。这个算法并不追求融合并创建所有特性的平均值——正如1877年弗朗西斯·高尔顿

（Francis Galton）在试验中发现的那样，当时他试图通过综合真实罪犯的照片来构建一个罪犯的原型形象时，却得到了一些与原始形象相去甚远的结果。将底片叠在一起，然后曝光、冲洗得到新的照片，高尔顿非常震惊地发现，他所使用的一组扭曲、丑陋的脸合成的照片中竟然呈现出一张英俊的面孔。数据科学家们想要创作出一幅可能被认为是伦勃朗的作品，就必须要有个更智能的计划。他们的算法必须创造出全新的眼睛、鼻子和嘴巴，就好像是通过伦勃朗的眼睛所看到的一样。

在创造完这些面部特征后，他们开始研究伦勃朗使用什么样的比例和关系将这些特征组织起来构成一张完整的面部。这也是令达·芬奇最为着迷的地方：他的速写本上满是对面部不同区域的相对位置的测量。有些人认为他是在利用黄金比例的数学概念来创造完美的面孔。伦勃朗并不痴迷于潜在的几何结构，他似乎对某种比例比较偏爱。

研究小组首先对平面图像进行分析，但绘画作品并不是简单的二维图像，画布上的颜料有助于画面产生三维效果。对于许多艺术家来说，这个特性和构图一样重要。想想凡·高创作油画时使用颜料的厚涂法，他像创作雕塑一样绘制一幅画。绘画中那些有质感的纹理是通过算法创作艺术的人经常忽略的东西，这种艺术通常只在屏幕上呈现，因此它受限于二维的“数字画布”。在伦勃朗晚期的作品中，分层绘画手法是一个



重要特征。研究小组意识到，现代3D打印机将给他们一个机会来分析和重构伦勃朗油画的分层结构特征。最终形成的3D打印画超过1.48亿像素，由13层紫外线固化数码喷墨制成。

巴斯·科斯滕（Bas Korsten）是参与该项目的创意合作伙伴之一，他承认，尽管这个设想非常简单和巧妙，但它的执行绝非如此。他说：“这是一段不断尝试和试错的旅程。我们有很多想法经过研究或测试都最终被证明是不可行的。我们曾考虑装配一个机械手臂来完成最终的绘画，但目前机械手臂只有9个自由关节，而伦勃朗的手至少有27个不同的部分可以独立运作，所以我们放弃了这种想法。”

科斯滕认为，最大的挑战是让“复活伦勃朗”背后的理念继续走下去。来自各方面的因素都在制造障碍，比如时间、预算、技术、舆论……但最重要的是，他们需要处理大量的数据。坚持不懈、不放弃，才是这个项目成功的唯一原因。

经过18个月的数据处理和500个小时的渲染，研究小组终于准备好向世界展示他们“复活伦勃朗”的成果了。2016年4月5日，在阿姆斯特丹，这幅油画向公众展出。展出的前几天，在Twitter上就有超过1000万条评论，说明它引起了公众广泛的关注，这个结果非常出人意料。不可否认，该作品再现了伦勃朗的绘画风格。如果让人们来定位这位“艺术家”，我相信大多数人会把他归入伦勃朗画派。他能续写伦勃朗的传奇吗？英国艺术评论家乔纳森·琼斯可不这么认为。

他在《卫报》上轻蔑而厌恶地写道：“这是对人性中所有创造性的可怕、无味、麻木和无情的嘲弄。”“在我们这个奇怪的时代，当最优秀的大脑将自己奉献给最愚蠢的‘挑战’时，当技术用于那些永远不应该使用的事物时，每个人都不得不为这样的结果鼓掌，因为我们如此推崇将一切数字化。这是多么可耻、可悲啊！”

琼斯觉得这幅画作没有抓住任何伦勃朗创作的特点。这无关风格和表面效果，而是伦勃朗通过绘画揭示、表达了他的内心世界，这样做也会揭示我们自己的内心世界。这是两个灵魂之间的碰撞。这幅人工智能的画作完全无法引发人们产生琼斯所描述的“伦勃朗的战栗”——那种置身于伦勃朗每一幅杰作前浑身起鸡皮疙瘩的感觉。

在他看来，只有一种方法能让这样的试验获得成功，那就是让人工智能体验瘟疫、贫困、衰老以及其他所有的人类的经历，因为正是这些

经历造就了伦勃朗的为人，也造就了他的艺术。

这样的评论公平吗？如果不是被事先告知这幅画是由计算机创作的，琼斯会做出相同的反应吗？艺术家的创作过程通常是一个密匣，算法给了我们新的工具来探索它内部的东西，并从中找到模式的痕迹。如果我们可以通过代码复制艺术家的作品，那么代码就揭示了创作的过程。这能否帮助我们识别出那些未被确认的已故大师的作品？能帮助我们剔除那些被错误地归属在大师名下的作品吗？

几十年来，围绕荷兰人威廉·范·德·沃姆（Willem van der Vorm）所收藏的《托比特和安娜》究竟是谁创作的，一直存在很多争议。这幅作品确实具备伦勃朗晚期作品的许多特征：集中的光线，画面笔触与质感纹路，局部的失焦，以及其他焦点清晰部分的选择。它的底部甚至有伦勃朗的签名，但许多人认为这是后来加上的，是伪造的。几十年来，这幅画都没有被归为伦勃朗的作品，而被认为是他的一个学生的作品。这一切在2010年发生了改变，研究伦勃朗的专家恩斯特·范·德·韦特林

（Ernst van de Wetering）将现代科学的力量运用于其上。红外线扫描和X射线分析技术在绘画上的应用，让我们可以看到隐藏在绘画表面之下的东西。比如，让我们了解艺术家在作品创作过程中，在最终定稿之前的整个过程中所做的每一次尝试。X光照片显示，这幅画表面之下最初还有着另外一幅画，但是被现在的画面所覆盖了：在底层的画面里多出了一扇窗户。韦特林这样说：“伦勃朗善于以这种方式不断地安排光线，尝试用不同的方式照亮人物。”微观化学分析也可以揭示出一个显著的特征，那就是油画必须在颜料还未干时完成绘制。根据韦特林多年的经验和对伦勃朗绘画风格的深刻了解，再加上这些新的科学技术的支持，他改变了对这幅画归属的看法。展出这幅画的博物馆得到这一结论时非常高兴，因为自己的馆藏又多了一幅伦勃朗的作品。尽管得到了科学证据的支持，但仍有一些评论家对这幅作品的作者归属持怀疑态度。

那么，韦特林是怎么看待这个“复活伦勃朗”试验的呢？起初，他听到这个消息时很是反感。当他最终见到这幅作品时，立即开始对这幅画的笔法进行仔细的审视，对微妙的不一致处进行评论，指出这幅画所使用的笔法是伦勃朗在1652年左右使用的，而其他的肖像画的笔法则更像是1632年使用的。即便是这样，研究小组依然感到相当宽慰：只有在精细到笔法这一层面的细节上，他们的项目才被发现存在不足。

对于微软而言，开展“复活伦勃朗”项目的动机很可能是商业原因，而非艺术。他们是想通过“再造”出令人信服的伪伦勃朗作品，来证明它

们所编写的代码有多么好。同样，AlphaGo与李世石对弈并胜出，与其说是为了发现新的、更具创造性的围棋玩法，倒不如说是为了向外界宣传DeepMind的人工智能实力。创造力应该与商业因素绝缘吗？凡·高生前只卖出了两幅画（尽管他会用画作换取食物和绘画用品），也许他只是希望过一种简朴的生活，但财富似乎并没有成为他创造的动力。然而，有证据表明，以金钱为奖励手段可以刺激人们的创造性产出（至少在低水平上）。

2007年，一个由美国心理学家组成的研究小组邀请了115名学生阅读一篇关于爆米花在平底锅里爆开的短篇故事。试验要求学生为这个故事起个名字。其中一半学生被告知：“我们将根据其他参与试验的学生为故事所起的名字来评判你所起的故事名字的创造性。如果你所起的故事名字比80%的人更具创造性，那么恭喜你，你出色地完成了这一试验。”另一半的学生被告知了同样的内容，但还被告知如果位于前20%，他们将因为自己的创造力得到10美元的奖励。果然，奖金的刺激带来了更多有创意的名字，包括“爆米花大混乱”<sup>[1]</sup>

（PANdemonium）、“天启”<sup>[2]</sup>（A-pop-calyapse Now），等等。

来自他人无论什么形式的反馈，是否就是创造的动力？难道人类不是在不断地进行发明创造以确保同类对我们怀有持续的兴趣吗？反馈已为新一代人工智能开启通向“整合”的大门。在机器学习中，反馈常常用于促使算法得到更好的结果。以DeepMind的雅达利游戏算法为例，“奖励冒险行为”规则（通过尝试不同的可能以寻求高分）促使算法突破了非激励算法的极限水平。

[1] 该名称中的PAN影射平底锅。——译者注

[2] 该名称由popcorn和apocalypse组合而成。——译者注

## 创造力竞争

除了证明可行性以外，“复活伦勃朗”项目是没有什么意义的。真正令人兴奋的新艺术能从代码中产生吗？罗格斯大学（Rutgers University）的艾哈迈德·埃尔加马尔（Ahmed Elgammal）猜想，如果把艺术创作变成一个竞争性的游戏，是否会促使计算机进入新的、更有趣的艺术领域？他的想法是创建一种算法，其首要任务是放弃已知的所有艺术风格，然后判断由算法自己所产生的艺术品是否具有与所有艺术风格都截然不同的特性，即真正独树一帜的艺术风格。这是一个典型的生成式对抗网络<sup>[1]</sup>的例子，这个概念由谷歌研究员伊恩·古德费罗（Ian Goodfellow）首先提出。该网络模型通过框架中（至少）两个模块，生成模型（generative model）和判别模型（discriminative model）的互相博弈学习，产生相当好的输出。埃尔加马尔希望开发出这样一种算法，并希望其创造性在世界范围内得到认可。

有一些证据表明，这种对抗性模型同样适用于人类创造力代码的引导。汤米·麦克休（Tommy McHugh）的神奇经历就证明了对抗性模型的力量。汤米是利物浦的一名快活的建筑工人，他结了婚，住在伯肯黑德的一所小房子里，除了在监狱里纹的文身，他对艺术毫无兴趣。但是在2001年，汤米中风了，中风之后，发生了神奇的事情：汤米突然有了创作的冲动，他开始写诗，买颜料和刷子，并开始在他房子的墙壁上创作艺术品。但问题是，他无法控制自己的这种冲动。他被这种冲动“劫持”，强迫自己在家里的每面墙上作画。

走进他的房子，就像走进了媚俗版本的西斯廷教堂，一切都被图画覆盖着。汤米的妻子拿创造力爆发的汤米毫无办法，只能对他听之任之。汤米停不下来，他一直不停地绘制新画来覆盖旧画。

他对我说：“我已经把整个房子画了5遍，地板、天花板、地毯……我只有累了睡觉的时候才会停止作画。如果允许，房子的外立面、树和人行道都会成为我作画的地方。”

他的作品都是杰作吗？答案是不尽然。但是，为什么汤米在中风后突然有了作画的冲动？他试图向我描述当这种创造的冲动出现时，在他的脑子里发生了什么：“在我的脑海里有一股闪电射向大脑的一侧，击中了其中的某个细胞，这个细胞激活了充满着泡泡的埃特纳火山<sup>[2]</sup>。

我脑海中每一个泡泡都与其他数十亿个泡泡相联系，当它们连锁爆发时，我所有的创造力也都跟着爆发了。”

神经科学家的研究发现，就像在Google Brain上运行的生成式对抗网络算法一样，人类大脑也有两个相互竞争的系统。一个系统是表现欲（生成模型），产生制造东西的冲动，是创造、表达的系统。另一个系统是抑制剂（判别模型），是对我们的想法产生怀疑、质疑和批评的系统。我们需要非常谨慎、仔细地平衡两者，这样才能有新的产出或是决定。一个创造性的思想需要与判别模型相平衡，通过反馈循环对创意进行评估，以便创意（生成模型）能够被修改并产出。

由此看来，汤米的中风使得他大脑中的判别模型失去了效用。虽说他所创造的可能并不那么伟大，但没有什么能让他停下来，因为在他的脑中只剩下了爆炸式的表现欲（生成模型），即想要创造出越来越多疯狂的图像和想法。

德国艺术家保罗·克利在他的教学素描本中表达了这种张力：“在创造艺术行为伊始，有了最初的想法之后不久，就出现了最初的反向运动，即接受性的最初运动。”这意味着，创作者对他创作的作品有质量的把控。

2012年，汤米死于癌症，但他对自己的遭遇并不感到遗憾，他这样说道：“我的两次中风让我经历了11年无与伦比、无人能及的精彩。”

埃尔加马尔的想法是编写代码来模拟艺术家潜意识中的生成模型和判别模型之间的反馈循环。首先，他需要建立一个判别模型，这个模型相当于艺术史专家，其评估生成结果。他与同事巴巴克·萨利赫（Babak Saleh）合作创建一种算法，该算法可以创作一幅从未面世的作品，并负责对作品的风格或作者进行归类。维基艺术（WikiArt）拥有1119位艺术家的81 449幅画作，在时间跨度上跨越了1500年，这可能是世界上最大的数字化图像数据库。是否可以创建一种算法，它可以通过维基艺术中的数据训练自己，随机抽取一幅画后，对其风格或艺术家进行分类？埃尔加马尔使用了部分数据作为算法的训练集，其余数据则用来测试算法。他应该怎样编写他的算法？哪些关键的因素可能有助于对这一庞大的艺术数据库进行分类？

想通过数据来对艺术家进行分类，我们需要一些具体的衡量标准。这一基本过程与声田公司<sup>[3]</sup>（Spotify）或奈飞公司背后的算法类似，但

你要寻找的不是个人品位，而是与众不同的特征。如果对数据库中的画作以两个不同的特点为目标进行扫描、测量、归类，那么每幅画作都可以用二维图上的一个点（横轴上的一点）来表示。那么，选择哪种特性来扫描、检测，会让你惊奇地发现毕加索的画作集中在横轴的一个区域，而凡·高的画作集中在另一个区域呢？

例如，以一幅画作中使用的黄色量为特征进行测量，你可能会看到毕加索的画作（标记为×）和凡·高的画作（标记为○）的分布如图8-1所示：

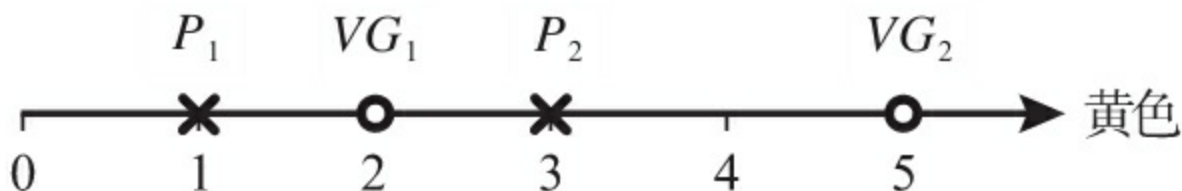


图 8-1

对单一特征进行测量并不能帮助我们判断画作归属。在上例中，有时候毕加索会用少量的黄色，比如 $P_1$  这幅画作对应的黄色刻度值为1。但 $P_2$  对应的黄色刻度值达到了3。而同是凡·高绘制的两幅画作 $VG_1$  和  $VG_2$  的黄色特征分布范围与毕加索的存在重合现象，故无法清楚地辨别两个人的作品。所以，单是测量黄色特征对我们是没有帮助的。

如果选择另一个特征来测量，比如画作中蓝色的使用量，会发生什么呢？我们在垂直的轴上以相同的画作为样本，标记出该特性的值，如图8-2所示。



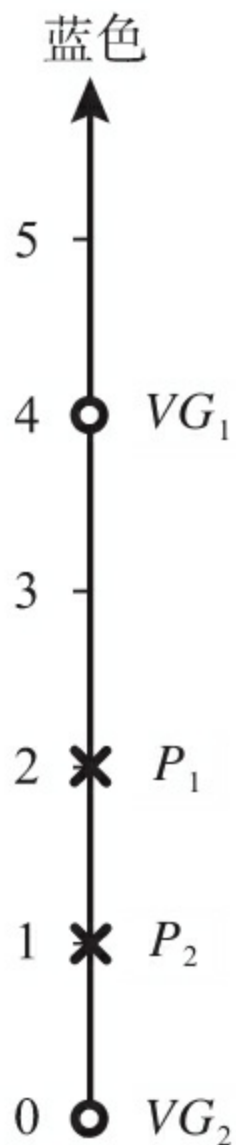


图 8-2

事实再一次证实蓝色这一单一的特征也不能真正帮助进行画作归属的判定，毕加索、凡·高的画作之间并没有明确的区分。但是，如果把这两个特征相结合，在平面直角坐标系里标出这些画会不会有所发现呢？

如图8-3所示，毕加索的画作 $P_1$ 对应平面中的点（1，2），凡·高的画作 $VG_1$ 对应平面中的点（2，4），依此类推可以找出所有作品的所在位置。将黄色和蓝色两个特征相结合，我们发现通过一条直线（图8-3中的虚线）可以将艺术家的作品分为两个不同的区域，毕加索的作品聚

集在线的左侧区域，而凡·高的作品聚集在线的右侧区域。

在算法学习了如何利用这两个特征来区分毕加索和凡·高的画作之后，拿出一幅新的绘画作品给算法，并要求它识别该画作的作者是凡·高还是毕加索时，算法将测量这两个属性，然后在平面直角坐标系中绘制出该画作的坐标，确定其所处位置，并根据这一结果判断出这幅画作的创作者。

在这个简单的例子中，我选择了颜色作为特征来区分艺术家，当然，也可以利用画作中其他众多的特征来进行分析。机器学习的强大功能是探索可能的度量空间，就像在以上示例中测量黄色和蓝色一样，其可以挑选出有助于区分艺术家特征的正确组合。显然，用来区分艺术家，只测量两个维度的特征是远远不够的，所以我们需要找到足够多的、可测量的、独有的特征。每一个新特征都有助于更好地标记艺术家和他们的风格，增加定位艺术家和他们风格的空间维度。最后，绘制出多维度的图表，而不是我们在上述的简单示例中用到的平面直角坐标系。

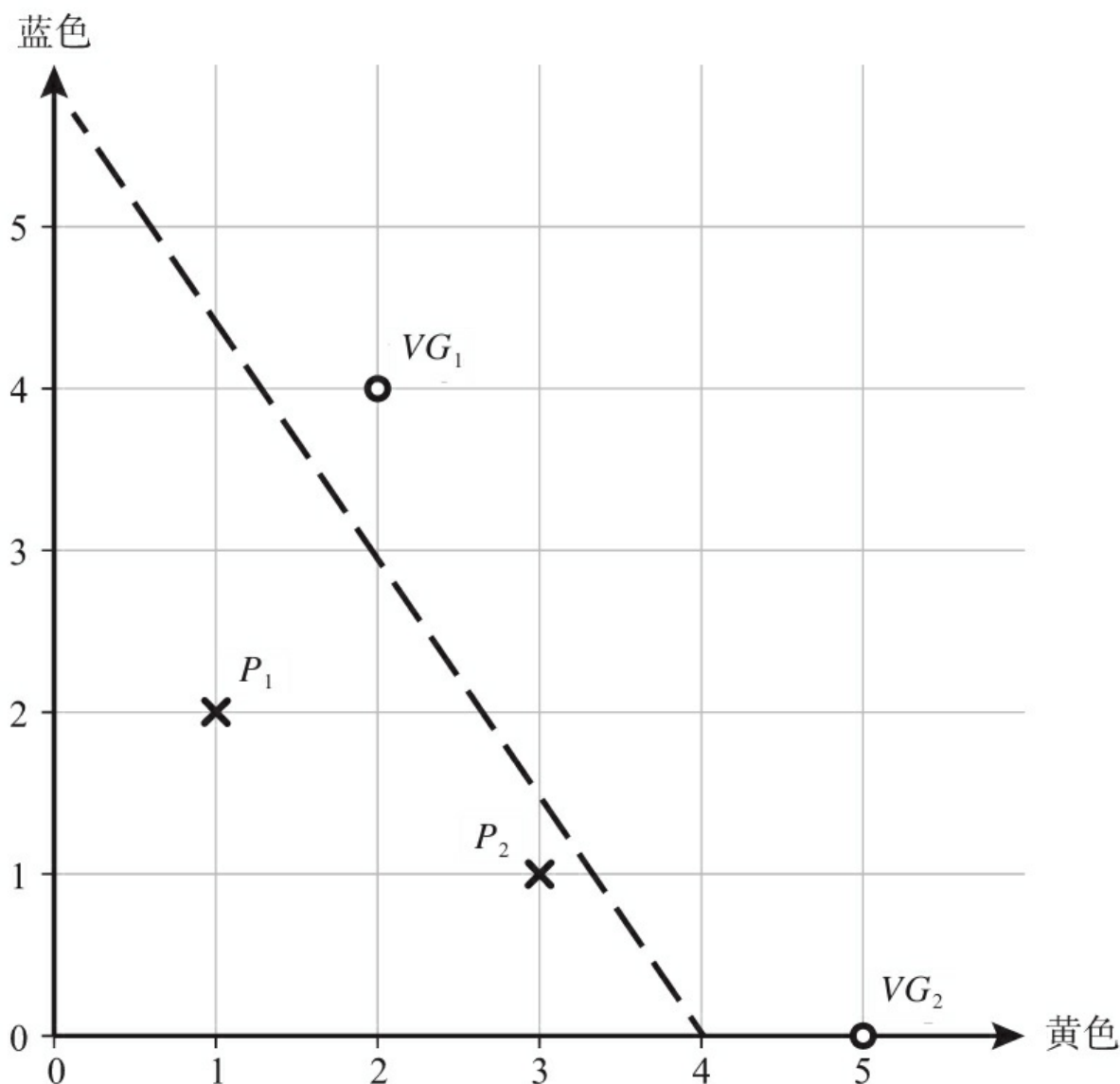


图 8-3

最近几年取得快速突破的最先进的计算机视觉技术，能够根据画作包含的视觉概念对其进行分类。这些概念被称为“classes”，包含了几乎所有的东西，从小的物质描述（如鸭子、飞盘、人、独轮手推车），到颜色变化，再到高层面描述（如死尸、水体、路面），等等。这种方法允许计算机视觉算法去分析图像，并生成一个对图像进行描述的 classes 列表。此列表像一张矢量图，会对图像进行定义，并且可以用于与其他以相同方式进行分析的图像进行比较。这些特征成为程序员改进他们算法的极佳依据。

在前文中，我们看到在平面直角坐标系中是如何可以将毕加索和凡

·高区分开来的。为了在现实的数据中精确区分不同的特征，该算法必须有效地对400种不同的特征进行分析，即在400个维度中分析画作。在对前所未见的画作进行测试后，该算法能够有超过50%的概率识别出这些作品的作者，但其很难分辨出莫奈和毕沙罗等19世纪末和20世纪初印象派画家画作之间的区别。可是，人类艺术家只需要做一点小小的研究就能明确区分他们的画作。19世纪末至20世纪初，莫奈和毕沙罗在法国非常活跃，并且都在巴黎Académie Suisse艺术学院进修，正是在那里，他们成为一对好朋友，经常会向对方分享自己在艺术创作上获得的经验。所以，他们的艺术作品有几分相似，就不足为奇了。

罗格斯大学的研究小组决定测试他们的算法能否识别出艺术史上代表新风格出现的作品或标志着转折的极富创造力的作品。这是因为一部分艺术家不断地挑战现有的艺术边界做出一些突破，而另一部分艺术家则创造出了全新的艺术风格。算法能识别出那些打破惯常规则，开创了一种新绘画风格的画作吗？算法是否可以识别出画作到底是立体主义的作品，还是巴洛克风格的作品？

该算法已经在多维度状态下分析、定位了所有的绘画，并把它们在高维图表中绘制成点。在添加画作的创作时间这个维度后，如果算法检测到某一幅画作在高维空间中沿着时间维度移动时发生了巨大的位置变化，那么这幅画作是否具有与艺术史专家公认的在艺术史上具有创造性革命意义的作品相一致的意义？

以毕加索的《亚威农的少女》为例，现在许多人认识到这幅画作打破了传统。然而，1916年《亚威农的少女》在巴黎首展时，正如你所预料的那样，这幅将带来审美史上重大变革的伟大画作，人们最初对它的态度是非常不友好的。一篇发表在报刊Le Cri de Paris上的评论文章指出：“立体主义者并不会停止挑衅，他们不停地对理智进行攻击。”但没过多久，这幅画作就被认为是艺术史上一个伟大的转折点。几十年后，正如《纽约时报》的艺术评论家所写的：“它一举将过去的艺术掀翻，无情地改变了我们这个时代的艺术。”令人兴奋的是，与同时代其他人的作品相比，算法在高维图表中发现了这幅画作的位置发生了巨大的偏差，因此将其高度评价为与以往任何画作有明显不同的画作。或许，就连《纽约时报》的艺术评论家也将被一种算法抢去风头，直至被替代。

罗格斯大学团队的算法判别模型就像一位艺术史专家，可以判断画作是否属于一种已被公认的艺术风格，并帮助判定画作是否开创了一种新的艺术风格。与之对应的是算法的生成模型，它的任务是创建一些被

认为是艺术品的，可被接受和认可的，不同以往的作品。埃尔加马尔为了理解艺术作品创新程度与美感唤醒之间的关系，深入地研究学习了丹尼尔·伯莱因（Daniel Ellis Berlyne）的研究成果。伯莱因对心理学的主要贡献体现在动机与唤醒、思维与心理美学，以及视觉艺术等领域。伯莱因特别关注美感的唤醒，他提出通过不同的刺激类型的特性，如新奇性、好奇心、复杂性、模糊性和费解性等，可以促使美感唤醒的产生。

[4] 该理论认为对唤醒的偏好水平是由个体决定的，而偏好水平又是影响个体行为的一个重要因素。一般来说，个体乐于接受中等水平的刺激，它会产生最佳唤醒水平。过低或过高的刺激水平都不为个体所喜好。每个个体都有不同的最佳唤醒水平，低于最佳唤醒水平时，个体将寻求刺激，而高于最佳唤醒水平时，个体将选择逃避刺激。

威廉·冯特 [5] 在这个问题上也有巨大贡献，他的冯特曲线就是描述这一现象的（见图8-4）。如果我们对艺术品的风格过于习惯，就会产生冷漠感和无聊感。这就是为什么艺术家们从来没有真正地将他们的作品风格固定下来，因为唤醒艺术家（最终是观众）美感体验的因素是个体化的、独特的。艺术家追求的是最大的美感体验价值，但矛盾点在于唤醒的刺激水平一定不可以过高，过高将使我们陷入冯特曲线的下降阶段。

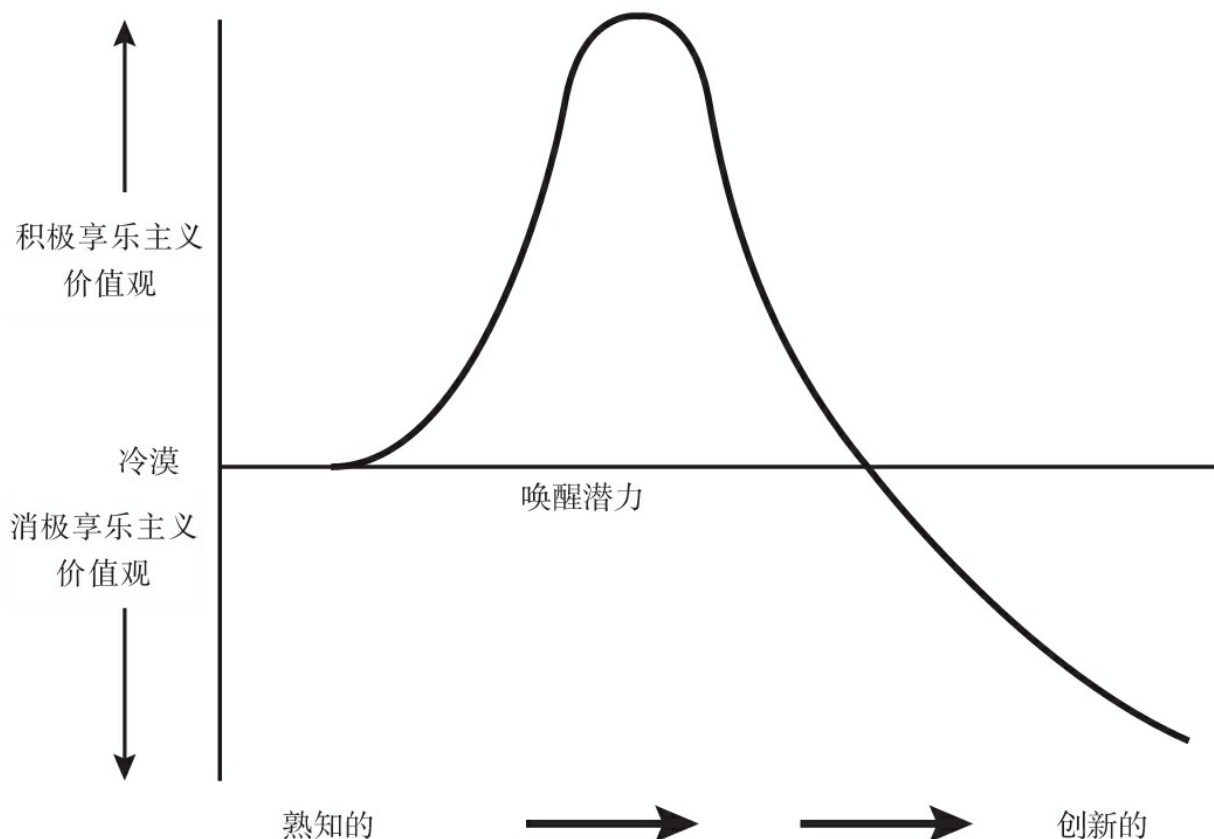


图 8-4

埃尔加马尔和他的团队为算法的生成模型编写了程序，以便算法创造出更接近于冯特曲线峰值的作品。这个试验的目的在于，在尽量不偏离艺术界公认的艺术品概念基础上使算法生成的作品最大限度地与现有艺术风格扩大差异。算法的判别模型将反馈结果给算法的生成模型，无论作品是过于守旧还是过于新潮、狂野，都不能被认定为艺术作品。每一次判别信息都会改进生成模型的参数，这就是机器学习的巧妙所在：算法会随着越来越多的数据而自我改进，从反馈中学习。人们希望生成模型能在判别模型不断地信息反馈中和自身不断地改进中，创造出落在冯特曲线峰值区的作品来。埃尔加马尔称这种算法为创意生成对抗网络（creative adversarial networks）。

那么，人们如何看待这些算法创作的作品呢？在2016年巴塞尔艺术博览会（Art Basel 2016）当代艺术展上，一些艺术爱好者展出了他们具有原创艺术风格的新作品。当他们把自己的作品拿来和埃尔加马尔团队创意生成对抗网络创作的“新艺术”作品进行比较时，发现由算法创造的艺术作品更具原创性、启发性，画面质量更加精良。



2018年10月，佳士得拍卖行（Christie's）成为首家出售由算法创作的艺术品的拍卖行，这或许意味着人工智能艺术越来越受到人们的重视。这幅画作由法国艺术团队Obvious创作，团队主要成员是三名青年（他们的教育背景均为商科或工科）。他们采用了古德费罗提出的生成式对抗网络（GANs），而没有采用埃尔加马尔提出的创意生成对抗网络（CANs）。这个创作团队采用了14世纪以来的超过15 000幅肖像画对他们的算法进行了训练（机器学习），机器根据训练指令创造出若干新作品，直到新作品成功通过判别模型做出的判断。

这幅肖像画的主角埃德蒙·贝拉米（Edmond Belamy）并非真人，而是由人工智能虚构出来的形象。在金色的画框中，是一个模糊的身影，他穿着清教徒式的黑色衣服，被巨大的阴影包裹着，神态朦胧。他也许来自法国，生活于某一个不可考证的历史时期，给人一种略显不安的感觉。这幅画面没有中心焦点，似乎被画者不愿出现在那里一样，整幅画作的抽象派气息浓郁。从艺术风格上讲，这幅非常现代的肖像画结合了18世纪的风格，又与英国艺术家格伦·布朗的风格相似，所以让人很难确定它的年代。画作右下角还有一个新奇的签名：不是艺术家的名字，而是一个数学方程式。

创作团队Obvious利用源代码及算法生成了虚构的贝拉米家族世系谱系列肖像画，《贝拉米肖像》是11幅系列肖像中的一幅。佳士得对这幅画作的描述是贝拉米伯爵的曾孙埃德蒙·贝拉米。《贝拉米伯爵》的肖像画于2018年2月被尼古拉·劳格罗·拉塞尔以1.2万美元收藏，而在佳士得的拍卖会上，伯爵曾孙的肖像以惊人的43.2万美元成交。贝拉米这个姓氏来自法文的“Bel Ami”，是好朋友的意思，“Goodfellow”也是这个意思，该团队把古德费罗的姓氏“翻译”成了法语用在了这里，是为了鸣谢和纪念他发明了生成式对抗网络。

大多数人类艺术家都会经历这样的过程：从过往的艺术家和他们的作品中学习，并利用所学到的知识，从模仿到推陈出新开创属于自己的艺术风格，创作出属于自己的艺术作品。对于艺术的理解，我们只能求助于人类的共同经验，毕竟在欣赏“新艺术”时，大多数观众看待它、理解它所依靠的知识体系和参照标准是观众自身的经历。巴塞尔艺术博览会上，无论是艺术家还是观众，就没有人从未接触过毕加索和蒙克的艺术风格。大多数创造力来自于这种改变现状的想法：创造一个脱胎于现状但又有别于现状的未来。这是一个进化模型，有趣的是，这正是算法所采用的。

你可能会觉得算法这种方法是可怕的操纵：把艺术变成一幅数字格式的画作，只为了找到能激发最大快乐价值的点。这听起来很可怕。难道只有表达了内心焦虑的才是伟大的艺术家吗？然而，这条通往艺术创造力的另类道路还是发挥了作用。这些生成式对抗网络算法可以带领我们进入一个新的领域，即使我们承认这是一门艺术，但作为人类，我们仍然对其束手无策，无法探索。总而言之，计算机代码具有揭示人类代码创造的艺术中未开发潜力的能力。

[1] 一种深度学习模型，是近年来复杂分布中无监督学习最具前景的方法之一。——译者注

[2] 意大利西西里岛东岸的一座活火山，海拔3200米以上，是欧洲海拔最高的活火山。埃特纳火山喷发状况十分活跃，自2007年以来发生了20余次喷发。——译者注

[3] 一个正版流媒体音乐服务平台。——译者注

[4] 伯莱因与唐纳德·赫布（Donald Olding Hebb, 1904.07.22—1985.08.20，加拿大心理学家，认知心理生理学的开创者）通过大量观察试验，提出了最佳唤醒（optimal arousal）理论。——译者注

[5] 1832.8.16—1920.8.31，德国生理学家、心理学家、哲学家，实验心理学之父。——译者注

## 算法如何思考

艺术可以做很多事，但对我来说艺术最好的地方在于，它给我提供了一个了解他人内心世界、思维方式的途径。这也许就是人工智能艺术的真正潜力所在，因为它有可能帮助人类理解隐藏在计算机底层代码中的本质。如果人工智能即将替代我们人类去完成许多事情，那么了解一下人工智能的“世界观”“价值观”“机生观”就是非常有必要和重要的了。

谷歌的一个团队一直尝试通过人工智能创造的艺术来更好地理解他们所创造的算法在视觉识别中的思维过程。正如我在第5章中所解释的，已经开发出来的用来区分猫和香蕉图像的算法识别图像依赖于算法关于图像的层级发问。该算法有效地设计并完成了由20个问题组成的层级结构以识别图片中的内容。

问题是，随着机器学习的演化，程序员越来越难以追踪算法用哪些特征来识别到底是猫还是香蕉。如果只是查看原始代码，很难对算法的工作原理进行逆向工程。算法可以针对图像提出数百万个不同的问题，我们很难看出算法为何选取这些问题，也很难明白问题之间的优先关系是怎样安排的。为了了解这个算法是如何运行的，谷歌的团队想出了一个很聪明的办法，那就是将算法的工作程序颠倒过来。他们出示一个随机的像素化图像给这个算法，并要求它增加或增强识别一些特征，而这些特征是算法认为会触发识别可识别特征的特征。他们希望通过这个过程能揭示出这个算法究竟在寻找什么。他们称这种反向算法为“深梦”（DeepDream）。

对我而言，“深梦”所创造的图像可能是我所见过的最有意义的人工智能艺术形式。这些图像让我们看到了视觉识别算法是如何看待世界的，而不是试图去再创作一幅“伦勃朗的作品”，也不是巴塞尔艺术博览会上与现代艺术家的比试。也许它在美学上的意义并不大，但对我来说，这就是艺术的全部：试着用另一种眼光去理解世界，用一种不同的方式去看世界。

“深梦”算法使用了一种人类看图像的方式：我们会从烤面包片的焦痕中突然看到一张人脸，或是在天空中的云彩里看到动物的样子。但其实，在面包片上、云彩中什么都没有。人类的大脑已经进化到对动物图像极其敏感，因为这是生存的关键。但这也意味着有时我们会有误判，

就像面包片上并不存在人脸一样。视觉识别算法也以类似的方式工作，其会试图寻找这种现象发生的模式并解释为什么会这样。“深梦”算法通过数千幅图像的训练，已经在压缩版的进化中学会了如何正确地发现、识别图像这一模式；它的“命”依赖于能否正确地识别图像。机器学习基本上是数字进化的一种形式，那么，“深梦”算法在数字化的灌木丛中看到了什么呢？

谷歌研究小组的发现令人相当惊讶。“深梦”算法像人一样无中生有地“看到”了海星和蚂蚁。似乎这个算法不仅有识别图像的能力，还有生成图像的能力。这不仅仅是一个有趣的游戏！这个试验为算法如何进行机器学习，提供了令人着迷的见解。在该算法中，哑铃的图像总是和一只握着它的手臂一起出现。很明显，该算法从人们举重的图像中“学到”了哑铃。正因为算法不懂人体解剖学，所以它不明白哑铃不是人体的延伸，它们可以独立存在。

你可以给算法一幅图像让它去识别，也可以邀请它玩一个我们都玩过的游戏：盯着天上的云，看看能在那些蓬松的云朵中发现隐藏着什么。算法能够在云朵中发现狗、鱼这些动物，或是发现狮身人面像、五爪金龙这些神话传说中的生物。《仿生人会梦见电子羊吗？》于1968年出版，作者是美国著名的科幻作家菲利普·迪克。广受好评、屡获大奖的电影《银翼杀手》就是由这部小说改编的。如果仿生人使用这些算法，他们就能拥有羊了——在算法生成的一张图像中，羊确实出现在了天空中。

算法将从人类的手中接手越来越多的决策，但问题在于，算法的机器学习导致了决策树的出现，而决策树对于人类来讲太复杂，我们很难对它进行分解。正因为如此，最终我们并不确定为什么算法会做出这样或那样的决定。这也是这种新型编程的局限之一。我们怎么才能确定算法做出的决定到底是一个错误，还是一个极具洞察力的建议呢？在AlphaGo与李世石的对阵里，第二盘棋的开局不久，AlphaGo就下出了让所有职业高手都无法理解一手棋。席尔瓦（AlphaGo技术负责人）称其为反人类的一手——第37手5路肩冲。席尔瓦解释道：“多数评论员都在第一时间批评这一步棋，从来没有人在这样的情况下走出如此一招。在胜负已定之后，一些专业人士重新思考这一步，他们改口称自己很可能也会走这一步。”越来越多的算法不仅仅用于游戏中，它们做出的决策影响着我们的生活，因此，在我们迈向一个自动化日益盛行的未来之际，任何能帮助我们理解算法的决策过程、决策策略以及价值判断的工

具都是至关重要的。

就计算机视觉算法而言，它们通过产生艺术作品给我们提供了一些线索，让我们能够了解算法是如何运作的。有时算法检测和选择的特征是我们能够识别出的，但有时我们似乎很难说出算法在图像识别中到底选择并检测了哪些特征。这门“艺术”给我们了一种抽象的洞察力去深入了解算法在决策树的特定层级上所做出的处理。我们可能正在逐步深入了解算法的深层潜意识，谷歌的程序员称这个过程为“开始主义”（inceptionism）。他们认为这些图像就像算法的梦境，因此为算法取名为“深梦”。算法所生成的图像对他们来说有一种疯狂的迷幻感，就好像算法施了魔法一样。通过在算法的输出上反复应用算法并在每次迭代后生成新作品，程序员可以产生无穷无尽的新感观。

我认为没有人会把“深梦”的作品当作优秀的艺术品（不管究竟它是什么），所以你绝对不会在伦敦Frieze艺术博览会或巴塞尔艺术博览会上看到它们。但是，“深梦”所做的仍然是让我们理解算法内部运作状态的一种重要的新方法。

## 算法就是一门艺术

这些新工具是否将视觉艺术带入了一个有趣的新领域呢？我决定回到汉斯·乌尔里希·奥布里斯特的画廊，与他聊聊人工智能在艺术世界中所扮演的角色，听听他的意见。但在去他的办公室之前，我决定先去看现在正在展出的艺术品。

我在画廊看到了由郑曦然（Ian Cheng）创作的《信仰之袋》（bag of beliefs, BOB），这是一种用代码创造的人工生命形态，采用了嵌合分支蛇的结构形式。展厅中共有六组《信仰之袋》，每一组都以相同的代码开始，但是由于这些生命形态的进化受到了参观者交互的影响，当我参观的时候，它们已朝不同的方向发展了。就像我的基因完全相同的双胞胎女儿，后天环境因素会给她们的发展带来很大的影响，使她们的性格截然不同。

就像里希特的《4900种色彩》一样，我觉得有必要解开《信仰之袋》内部深处的代码。但这是另一种代码，一种更难逆向工程的代码。这也许就是为什么它能成功地比人们所预期的更长久地吸引人们的注意力——它是在与参观者互动的基础上学习和发展的。

《信仰之袋》通过与智能手机的互动了解参观者的情绪状态。郑曦然对原创和起源的问题很感兴趣，他想知道：“艺术是如何在其意义上被创作出来的，却又活在作者之外并自我变异的呢？”答案是创建一个系统，并允许其交互的内容在没有限制条件的基础上发展和变异。《信仰之袋》与参观者的交互意味着将郑曦然（作者）排除在外，代码会从它遇到的新参数中获得信息。

通常，我们人类会给自己无法理解的事物指定一个替代物来对它做出反应。当我们还不理解地震或火山爆发的原因时，我们创造了神，来为这些难以捉摸的力量负责。《信仰之袋》的核心算法也激发了参观者同样的反应，哲学家丹尼尔·丹尼特（Daniel Dennett）将这种现象称为意向立场。

正如汉斯·乌尔里希告诉我的：

在通常情况下，画廊的意见簿上的留言都是抱怨画廊太热，或者投诉椅子不够多，或者是关于他们喜不喜欢格雷森·佩里之类的。但是关



于《信仰之袋》的留言是：“它为什么不喜欢我？我为它感到难过。它不理我。”“它太可爱了！”能得到这样的留言真是太神奇了。

一天晚上，汉斯接到了画廊安保人员的电话，这时他已经出国旅行一周多了。安保人员告诉他：“今天凌晨3点，画廊突然被光照亮，不是火警，而是《信仰之袋》发出的光，这件作品活了，不受控制了！”尽管他最初设定《信仰之袋》从上午10点一直运行到下午6点，和画廊营业时间一样，但是该作品自己决定了唤醒时间。我们无法理解《信仰之袋》为何会在午夜醒来，这让我们觉得它可能获得了某些代理权限。正是这种令人费解的算法能力，给了由算法导致末日的影视作品和小说以素材和想法。

汉斯对我说：“大多数挂在画廊里的艺术品都是静态的、凝固的，是不会随着时间而改变的实物。若是视频艺术，也总有开始和结束。过去画廊里的任何视频都必须循环播放，而在你看了20遍之后，最终会审美疲劳觉得无聊。使用人工智能则打破了这种规则，它不需要循环地使用原素材。这种艺术作品不断演进、永不重复的开放性，是艺术界的新事物。”

《信仰之袋》背后的代码与波洛克滴画背后的模拟代码有一些共同之处。它基于受环境影响的确定性混沌方程，让参观者可以影响算法的输出结果。混沌带来了不可预测性，利用混沌的代码可以满足“创造性”所要求的新奇和惊喜。虽然混沌是确定性的，但如果我们想要打破从编码者到创造者之间的壁垒，它可能仍然是我们所能期望的最好途径。

乔纳森·琼斯在《卫报》的评论里只给了《信仰之袋》一颗星，他这样写道：“它们只是聪明的实验室模型，是没有灵魂的傀儡。艺术永远是以人为核心的，没有这一点它什么也不是。郑曦然显然忘记了这一点，他的作品就是一个无聊的电子产品。”尽管琼斯认定机器没有灵魂，但随着进入未来，我们将越来越需要利用画廊这种地方作为桥头堡，便于我们第一时间知悉第一个人工智能的灵魂何时出现。

汉斯认为对于人工智能的灵魂来说，艺术是最好的早期预警系统之一。考虑到社会上关于人工智能所扮演角色的辩论，在汉斯看来，人工智能在画廊中占据一席之地似乎是当务之急。当今，许多算法的应用是隐藏不可见的，我们在不知情的情况下被算法影响和操纵着。使用艺术将算法可视化，可以帮助我们更有意识地主动解释和指导这些算法。视

觉艺术家是人群和代码之间强有力的媒介，画廊展出的人工智能作品就是艺术。

汉斯对我说：“艺术家是化无形为有形的专家。那么，人工智能可以创造伟大的艺术，而不只是成为艺术品吗？虽说迄今为止还没有任何一件伟大的艺术品是由机器创造的，但是永远不要把话说得太绝对。我对未来持谨慎态度。我们永远不能排除机器可以创造出伟大作品的可能。当围棋选手们说机器永远赢不了我们时，AlphaGo证明他们错了。我是一名策展人，但我绝不会自傲地说，一台机器永远无法策划出更好的展览.....”

我能看到他的神经细胞开始活跃起来。他又说道：“总有一天人工智能会去策划并运营展览，这将是一个危险但有趣的试验。”

## 第9章 数学的艺术

亨利·庞加莱（Henri Poincaré）

灵光乍现，从来都是厚积薄发。

我13岁时萌生了成为数学家的想法，那时的数学老师向我推荐了几本书。在那个时候，我真的不知道成为一名数学家意味着什么，但他向我推荐的一本书给了我答案：数学不仅仅是计算的事。剑桥数学家哈代（G.H.Hardy）的《一个数学家的辩白》给我种下了成为一名数学家的种子。

这本书是一个启示，哈代想要从中传达数学的含义是：

数学家就像画家或诗人一样，都是形式的缔造者。如果说数学的形式比其他的更持久，那是因为数学的形式是由思想构成的。数学家的算式就像画家的画或诗人的诗，必须是美的。思想就如同色彩或是文字，必须以和谐的方式结合在一起。美是首要的，因为在数学的世界里，没有丑陋数学的容身之所。

我从没有觉得数学是一门创造性的学科，但通过阅读哈代的书，我发现对于数学来说，美学的敏感性和思想的逻辑正确性一样重要。

为什么我的老师会觉得我能成为一个数学家，而不是一个画家或是一个诗人呢？多年后，我这样问他，他答道：“在我观察到你很喜欢有创意的绘画时，我就知道你对抽象思维的反应是别人所不及的。”这是一次经过判断分析后完美的干预，使我对一门学科的渴望得到了满足。数学——创造性思维和逻辑绝对确定性的完美融合，正对我的胃口。

多年来，我一直相信，数学的创造性使它不会被计算机自动化轻易替代。但现在，算法正在“再制”伦勃朗那样的肖像画，并在巴塞尔艺术博览会上展出与人类绘制的画作相媲美的艺术作品。算法可以很快地达成黎曼<sup>[1]</sup>（Riemann）的数学成就吗？或是和发表在《美国数学学会期刊》上的论文竞争吗？我是否应该开始找其他的工作了？

哈代讨论数学就像玩游戏一样。他喜欢用下国际象棋来做比喻，但

自从计算机下国际象棋赢了人类以后，我一直拿围棋当我的挡箭牌，好让那些想以电脑干活比我又快又多为理由开除我的人闭上嘴。数学有关直觉，即使我不确定为什么我有那种感觉，我也能感觉到该以怎样的逻辑去探索未知。但当DeepMind的算法发现怎样以非常相似的方式去做某些事的时候，它引发了一场“生存危机”。

算法能玩数学家的娱乐项目——围棋，那它能证明定理吗？作为数学家，我最大的成就之一就是在《数学年刊》上发表了一个定理。安德鲁·怀尔斯（Andrew Wiles）在这本期刊上发表了他对费马大定理的彻底证明。那么，要等多久我们才能在《数学年刊》上看到由算法撰写的论文呢？

竞赛中，理解规则是很重要的。我应和计算机竞赛什么？肯定不会是让我坐在办公桌前做大量的计算。如果是那样的话，几年前计算机就该让我失业了。那么，数学家到底在做什么呢？

[1] 1826—1866，德国著名数学家。黎曼的工作直接影响了19世纪中后叶数学的发展，在其影响下，数学许多分支取得了辉煌的成就。黎曼对偏微分方程及其在物理学中的应用有重大贡献。对物理学本身，如热学、电磁非超距作用和激波理论等，黎曼也做出了重要的贡献。黎曼首先提出用复变函数论研究数论的新思想和新方法，开创了解析数论的新时代，并对单复变函数论的发展有深刻的影响。他是世界数学史上最具独创精神的数学家之一，开创了黎曼几何，并且给后来爱因斯坦的广义相对论提供了数学基础。——译者注

## 数学证明的游戏

如果你看到了一篇关于数学的新闻报道，它大概率是这样的内容：一位数学家“证明”了一些伟大而杰出的猜想。1995年，报纸上盈千累万的头条都是关于安德鲁·怀尔斯对费马大定理的彻底证明。2006年，特立独行的俄罗斯数学家格里戈里·佩雷尔曼（Grigori Perelman）证明了数学中一个重要的未解决的问题——庞加莱猜想

（Poincaré conjecture），这使他获得了赢得百万美元奖励的权利。还有6个“千禧年大奖难题”，它们向数学家发起了挑战：要想证明自己学科的猜想，即使有直觉也依然棘手。

数学家工作的核心是证明。公理是关于数字和几何的不言自明的真理，证明就是从公理开始的逻辑论证。通过分析公理，我们可以重新组合出关于数字和几何确切的新的表达形式。然后，这些新发现可以构成新证明的基础，而新证明反过来又将引导我们发现公理的更多逻辑结果。数学的发展就像一个有生命的生物体，从先前存在的形式向外不断延伸开来。

人们常把数学证明比作下国际象棋或围棋。公理是棋盘上棋子的起始位置，逻辑推理规则是决定棋子如何运动的参数，证明是棋子一步一步的运动轨迹。在下国际象棋时，每一步棋都可能有成千上万种可能。例如，开局四步棋之后（黑白各两步），在棋盘上，棋子的分布就已经有71 852种可能了。通常，你不需要走几步棋就能达到这样的效果。对于围棋来说，棋子分布可能性的数量更甚。

如果我把棋子随机放在棋盘上，你可能会问，有没有可能从初始状态把棋一步一步走成这样？换句话说，随机摆在棋盘上的棋子位置，按照围棋或是国际象棋的规则是可能的吗？这类似于数学中的猜想，例如费马大定理。费马断言当整数 $n > 2$ 时，关于 $x$ 、 $y$ 、 $z$ 的方程 $x^n + y^n = z^n$ 没有正整数解。这本身就是一个猜想。数学家所面临的挑战是需要证明得到这样的结果是否符合数学本身的逻辑。费马就是这样把棋子摆在棋盘上，然后说：“我相信你一定能按照棋的规则，把棋一步步走成这样。哈哈哈哈！”安德鲁·怀尔斯和其他为证明费马大定理而努力工作的数学家，就这样确定了“棋子”一系列的移动，最后完成了费马大定理指定的排列方式。

数学界的艺术之一就是找出这些猜想目标。许多数学家认为，提出正确的猜想比埋头苦算更重要。要发现暗藏在数字里的真相，需要对数学有异常灵敏的嗅觉。这往往就是数学家最具创造性和可以发挥高深莫测技能的地方。数学家只有一辈子都沉浸在数学的世界里，才可能获得关于数学猜想的灵敏嗅觉。这通常是一种不需要解释的直觉和预感，是所有人梦寐以求的东西。

这就是计算机很难对猜想计算成功的原因之一。自上而下的算法像是一个醉汉在黑暗中跌跌撞撞：它有可能会随机地溜达到一个“有趣的地方”（奇异点），但大多数时候，它的行动没有重点、没有方向，毫无价值。但是，如果算法基于人类数学家的经验进行学习，这种自下而上的结构能否使算法发展出一种对奇异点的直觉呢？

数学家们是如何建立起这样一种对奇异点的直觉的？这种直觉通常不是巧合——在你脑海里往往有众多案例支撑，或者说应该是存在某种模式的。但是，这种直觉往往稍纵即逝，所以证明出一个猜想是如此的难得和重要。有时，需要数年才能发现一种模式是错误的。我在自己的工作中对一个模式做了一个猜想，一个研究生花了十年的时间才证明了它是错误的。

关于错误猜想，我最喜欢的一个例子是19世纪伟大的数学家卡尔·弗里德里希·高斯（Carl Friedrich Gauss）对质数的猜想。高斯认为  $\text{Li}(x) - \pi(x)$  的值总是正的，而且是递增的。所有的证据都表明高斯是对的。如果让一台计算机来解决这个问题，它将产生支持高斯猜想的数据。然而，1914年李特尔伍德从理论上证明了事实正好相反（即存在  $\text{Li}(x)$  小于  $\pi(x)$ ）。高斯的猜想是错误的，但证明他错误的这个数字大得惊人（注：李特尔伍德的学生塞缪尔·斯克维斯（Samuel Skewes）首次证明，如果黎曼猜想成立的话，第一个李特尔伍德反例值一定小于这样一个数，我们称之为斯克维斯数，其表示成简单的科学计数法是： $10^{100\,000\,000\,000\,000\,000\,000\,000\,000\,000\,000\,000\,000\,000\,000}$ 。——译者注），比宇宙中原子的数量还多（注：我们整个可观测宇宙的原子数不过是  $10^{80}$ 。——译者注）（即便这样，我们也无法接近这个猜想的崩溃点）。

这就是所有猜想所面临的问题：我们无法证明它们是真的，还是我们的直觉和现有的数据将我们引入了歧途。为了将那些未经证明的猜想与现已证明的定理联系起来，我们痴迷于尝试建立起一系列数学运算。

究竟是什么驱使人类去证明？人类创造数学的动机是什么？编写算



法来给数学家制造更多的挑战，这会成为我们探索数学领域的新动力吗？数学的起源可以追溯到人类试图理解自己所生活的环境，预测接下来会发生什么，从而使我们更加适应环境，并选择对我们有利的事物。可以说，数学是人类的一种生存行为（我在故我思）。

## 数学的起源

数学家总是被大家误解。可能大多数人都会这样想象：作为一个数学家，我就必须坐在牛津大学的办公室里，计算着一个有很多很多位小数的数，或者直接对六位数相乘进行口算。诚如哈代所言，数学家本质上是一位规律的探索者和发现者，而数学是发现和解释规律的科学。

正是这种发现规律的能力让人类在与自然世界的谈判中占据了优势，也正是因为它，让我们能够规划未来。人类非常善于发现这些规律，因为那些错过规律的物种没有能存活下来。当我遇到有人宣称（这种事经常发生）“我没有数学的头脑”时，我就会反驳道：“事实上我们都进化出了数学的头脑，因为我们的大脑善于发现规律。”有时，大脑的工作方法太先进了，会把图案解读成并不存在的数据，就像许多观众看到里希特《4900种色彩》系列绘画作品时感受到的一样。

我发现，对规律最原始的识别体现在一些最原始的绘画艺术中。拉斯科的洞穴壁画描绘了动物奔跑的精美画面，在这些静止的画面中，人们惊奇地发现了成群结队跑动的野牛。为什么这位艺术家要绘制这些图像，他是以什么样的身份绘制这些图像的？数学家、画家、史学家，抑或其他？

除了这些图像本身，我认为在表象之下还有一些最早的有关数学的记录。壁画上有这样的一些内容：昴宿星团，这是离我们最近也是最亮的几个疏散星团之一，在北半球晴朗的夜空中用肉眼就可以看到它；13个连成一串的圆点，在第13个圆点上方有一只拥有巨大鹿角的牡鹿；连成一串的26个圆点，在最后一个圆点上方是一匹怀孕的马。这些圆点代表了什么？有一种推测是这样的：每个圆点代表一个月的 $\frac{1}{4}$ （大约一周）。13周大约是一年的 $\frac{1}{4}$ ，那么，也许这些点是在描绘一个季节。处于北半球，当昴宿星团黄昏时就出现在天顶的这个季节（秋季9~11月），是狩猎牡鹿的好时机——在这个时候，它们正处于发情期，是脆弱的。

为了传递这些信息，必须有人发现并指出，动物的一种行为模式似乎每年都会重复出现，而这种行为模式与月相的变化一致。人们认识这种模式的动机显然是出于实际的需求，即推动发现的是实际效用。

在这里我们看到了数学的第一要素：数字的概念。能够精确地计算

出数字的意义对许多动物的生存至关重要，它会告诉动物在面对对手时，是该战斗还是逃走。通过对刚孵化的小鸡进行的复杂试验，证实了对数字认知的复杂能力是大脑固有的，与生俱来的：小鸡可以判断出5个比2个多，而比8个少。

但是，给这些数字命名并用符号表示是人类特有的能力。人类数学发展史的一部分是以一种“聪明的方法”识别并命名数字。古代玛雅人用点来表示数字，有多少就点多少点。但当数量变多的时候，这种方法就不是那么好用了，因为一眼看过去你很难区分到底是6个点还是5个点。所以，有人想出了一个“聪明的方法”：在4个点之间画一条线来表示5。就像外国电影中，囚犯在监狱的墙上画线计算日子一样。<sup>[1]</sup>

罗马人使用了一种新的计数体系，他们赋予了数字新的名字：X代表10，C代表100，M代表1000。古埃及人则使用新的象形文字来表示数字末尾的零：马蹄形代表10，一卷绳子代表100，一株荷花代表1000.....

但是，这些系统很快就失效了，因为我们使用的数字进入了数百万甚至数十亿的级别。每个新的巨大的数字都需要新的符号来表示。

玛雅人进行着复杂的天文学研究，他们需要大量的数字来记录大量的数据。他们想出了一个聪明的办法来解决罗马数字表达的问题，这就是我们今天用来记录巨大数字的系统。在我们的十进制系统中，数字的表达对应的是10的不同次方（幂）。以123为例，它表示有3个单元，1个100（ $10^2$ ）和2个10（ $10^1$ ）。不超过10的计数没有什么特别之处，我们可以用手指计数到10。事实上，玛雅人使用的是二十进制，数字的表达对应的是20的不同次方（幂）。比如玛雅数字中的123，表示有3个单元，分别是1个400（ $20^2$ ）、2个20（ $20^1$ ）和3个1（ $20^0$ ）。所以，它换算到十进制中对应的数字是443。

玛雅人并不是第一个想出用幂来表示数字这个聪明办法的，只不过其他文明使用了十进制或其他进制，而他们使用了二十进制。4000年前，古巴比伦人提出了独特的计数体系：他们没有采用玛雅人的二十进制，也没有采用我们现在使用的十进制，而是采用了六十进制，开创了一个新的体系。60可以被2、3、4、5、6、10、12、15、20、30整除，这种高可除性使它成了这个计数体系的基础，同时有利于进行高速有效的计算。

在实际应用中，计数体系的有效性、实用性、必要性等因素决定着它是否能够存在下去。六十进制的影响之一表现在今时今日我们记录时间的方式上：一小时是60分钟，一分钟是60秒。拿破仑曾经尝试让计量局使用十进制来计量时间，但很幸运的是，这种方法从未被真正大面积使用过。

在古巴比伦人留下的楔形文字泥板上，我们首次看到了数字与我们周围世界关系的数学分析。伴随着幼发拉底河沿岸城邦的发展，进行城市建设、征税、经商都需要数学作为计算的工具——更复杂的数学诞生了。出土的楔形文字泥板向我们展示了其官方的计划表，例如工人的数量、运河修建的工期、工程人工成本汇总，等等。这一时期的数学并没有什么特别有挑战性或有趣的地方，但很明显，它启发了一些运用数学的人去思考数学的其他可能性。

他们开始寻找一些简便的方法来帮助计算。从出土的泥板上我们发现了像我们所用的数学用表上的一些数据，例如数字的平方等。正因为有人发现了数字相乘与相加之间的关系，使得这些泥板成了比较大的数字相乘运算的辅助工具。例如这样一组代数关系：

$$A \times B = [(A + B)^2 - (A - B)^2] / 4$$

古巴比伦人意识到，他们可以利用他们的数学用表来便捷快速地算出A×B的答案。首先计算出A+B的值，然后查表得到(A+B)<sup>2</sup>的值，再减去(A-B)<sup>2</sup>的值，最后把得数除以4，就可以得到A×B的答案。这作为非常早期的算法，简直太让人兴奋了。这种算法将两个数字之间的相乘运算简化为相加运算，只要数字的大小不超过他们“泥板数学用表”的范围，使用它就可以很简便地得到答案。

虽然古巴比伦人以代数的方法来思考运用数字，但是他们并没有记录下为什么这种方法或算法总是能给出正确的答案，也就是说他们并没有将成果理论化，只是在运用而已。理论总是出现在实践之后。直到公元9世纪，巴格达智慧馆的波斯学者才发明了代数语言，上述方程才有可能被写下来，这时时间已经过去了几千年。智慧馆的图书馆长和天文台长，数学家、天文学家、“代数之父”花拉子密创立了代数这门学科，尽管最早使用代数的是古巴比伦人。

数字与数字之间的数学关系被更进一步有效地利用和加强，使得计

算的速度得到大幅提升。这种数学上的进步对于商业和建筑业的发展有巨大的推动作用。从发现问题到解决问题，这个过程看似注重实用性，但如果仔细思考，实际上它更倾向于从实践走向理论，更像是古代数字的使用者对于数学理论的研究，而非体力劳动者的所思所想。例如，有这样一个问题：某农民有一块面积为60个单位的田地，这块地的一边比另一边长7个单位，那么最短的边的长度是多少？

问题是我们不知道边的长度，怎么知道面积？对我来说，这更像一个填词游戏：出题人会对这个词进行相当复杂的描述，解题人需要正确补全信息才能解出这道题。对于上题，那条较短的边的长度我们设为X，那么较长的边的长度即为X+7，整块田地的面积为60，也就意味着两个边长的乘积为60。我们可以得到这样一个方程式：

$$\begin{aligned} X \times (X+7) &= 60 \\ \text{或 } X^2 + 7X - 60 &= 0 \end{aligned}$$

像这样的一元二次的方程，学校里的学生必须要学习如何解。当然，你可以心中暗自不满：“古巴比伦人为什么要发明这东西？”但是，我们应要感谢他们发明了解开这个方程的方法。

对我的专业来说，这是一个重要的转折点。为何会有人想到这些问题并去思考怎样解决？在日常生活中并不会出现这样的问题，我们为什么还要求学生如何解这样的题？可能这个农民之前已经计算过面积并把它写下来了，但后来他忘记了边长到底有多长，但是为什么他会知道长边比短边长7个单位，而又不知道短边有多长呢？这一切都太刻意、太做作了，这从来就不是一个真正的实际问题。所以，我得出结论：这道数学题的出现仅仅是为了好玩！

我们的大脑非常享受解决问题找到答案时的快感，这种快感是多巴胺或肾上腺素带给我们的。可以这样说，生物学和化学在推动着数学的发展。现在我们知道，就算没有数学题让我们解决，只要打上一针，我们也可以感受到这样的快感。计算机不是生物体，它无法产生多巴胺和肾上腺素，所以它也无法体验这种快感。那么，它还会为了好玩这样的动机去解决数学问题吗？

诚然，我们可以围绕善于数学运算的人在进化上更具优势这个命题而争论。事实上，这也是我们为何仍坚持在学校教学生如何解一元二次方程的最佳理由。解决这样的数学题，需要严密的逻辑思维、对抽象的

分析能力等，这些能力有机地结合起来才最终促成了解得答案这样的结果。这些能力投射在生活中就是解决实际问题的能力。

也许，解决一个数学难题带给我们生理上的快感是区分人类创造力和机器创造力的关键。计算机在构造上与大脑相似，我们或许可以创建一个由数字神经元组成的抽象网络，通过控制神经元之间的连接关系来模拟大脑的工作机制。但是，在这个系统中没有多巴胺和肾上腺素，计算机无法体会生物化学带来的生理上的快感。缺失了这种快感，计算机会缺乏创造性思维的动机和动力吗？

古巴比伦人的数学，是对特定的事例进行运算。他们发现的方法是用来解决某些特定的问题，他们只是发现了这样运算是正确的，但没有去解释为什么这些方法总是有效。直到几千年以后，数学才发展出证明这一概念。

[1] 中国人用画“正”字来计数“5”。——译者注



## 证明的起源

这种数学证明游戏的起源可以追溯到古希腊人，他们发现运用逻辑论证可以获得关于数字和几何颠扑不破的真理。证明是数学的本质特征，是数学家赌上自己的名誉在寻找的“圣杯”：要想赢得百万美元的奖金，你必须至少证明出“千禧年大奖难题”中的一个；要想赢得菲尔兹奖<sup>[1]</sup>，你必须能拿出一个让数学家同行印象深刻的证明。欧几里得的《几何原本》成了后世证明的范式，也为证明制定了规则。

欧几里得的巨大历史功绩不仅在于建立了一种几何学，更在于首创了一种科研方法。这方法所授益于后人的，甚至超过了几何学本身。欧几里得是第一个将亚里士多德用三段论形式表述的演绎法用于构建实际知识体系的人。欧几里得的几何学是一个严密的演绎体系，它从为数不多的公理出发，推导出众多的定理，再用这些定理去解决实际问题。比起欧几里得几何学中的几何知识而言，它所蕴含的方法论意义更重大。现在，再一次以国际象棋为例，来解释数学证明的运行方式：国际象棋的棋子在开局时的摆放位置相当于数学证明中的公理。公理是大家公认的、接受的、显而易见的关于数学和几何的事实。欧几里得的证明就是从这些公理开始。欧几里得本人对几何学的实际应用并不关心，他关心的是他的几何体系内在逻辑的严密性，譬如：任意一点到另外任意一点可以画直线；如果 $A=B$ ， $B=C$ ，那么 $A=C$ ；以任意点为圆心，任意长度为半径均可以画圆； $A+B=B+A$ 。

现在我们知道了如何摆放棋子，接下来我们该学习如何下棋了。棋子如何移动受到某些规则的限制，即这些规则决定了棋子的移动，而逻辑推理规则也允许我们根据迄今所知的事实来写下公理和真理。命题演算分离规则（modus ponens）是一种推演规则，指在命题演算和谓词演算形式的公理系统中广泛使用的推演规则，此规则的符号表示为 $A \rightarrow B$ ，即从 $A$ 可推演出 $B$ 。此规则的逻辑意义是，如果一个蕴含式及其前件均为逻辑真的，则它的后件也是逻辑真的。分离规则保持了永真性，即如果 $A$ 和 $A \rightarrow B$ 是永真的，则 $B$ 也是永真的，反之亦然。此规则的补充规则规定， $A \rightarrow B$ 为真时，若 $B$ 为假，则 $A$ 亦为假。

现在，让我们推理一下2的平方根无法用分数来表示。

分数是一个整数和另一个整数的不等于整数的比，其表示一个数是

另一个数的几分之几，或一个事件与所有事件的比例。分子在上，分母在下。无理数，也称为无限不循环小数，不能写作两个整数之比。若将它写成小数形式，小数点之后的数字有无限多个，并且不会循环。因为2的平方根是一个无理数，而分数属于有理数，所以2的平方根无法写成分数。

对我来说，一个游戏的结构是否良好且令人满意，在于它的规则是否易于理解和实现，同时，在规则范围内该游戏能提供给人极其丰富和多样化的操作空间。“井字棋”很容易理解和上手，但它很快就会让你觉得乏味，因为它可以拓展的可能性太有限了。在国际象棋和围棋中，就不会出现这样的问题。

玩国际象棋或围棋等游戏与玩数学证明游戏之间的一个重要区别是，数学家不必每次都把棋子归回原位：你可以从之前的任何一个时刻开始，以它为基础继续。在某种程度上，前辈数学家已经建立的和扩展出的公理大大扩展了你可以开始研究的范围，同时也为你提供了大量的可使用的招数。

惊为天人的是，我们赋予了符号和文字以意义：在纸上一划，这就是直线；用X来表示一个计数或测量某物的数字。那么，计算机怎样知道我们在表达什么呢？最美妙的在于我们在尝试探知数字和几何是如何运行的，我们试图从宏观角度去象征性地观察整个“游戏”。事实上，如果公理为真，我们赋予任何符号任何意义，都将引发一场“游戏”（推理与证明），推理与证明会帮我们找出答案。这意味着，计算机可以在不需要真正了解符号含义的情况下进行推理与证明。19世纪数学家戴维·希尔伯特<sup>[2]</sup>（David Hilbert）这样说道：“我们可以用桌子、椅子、啤酒杯来代替点、线、面。”当然，他的意思不是说几何学研究桌子、椅子、啤酒杯，而是在几何学中，点、线、面的直观意义要被抛弃，人们应该研究的只是它们之间的关系，而关系由公理来体现。几何学是对空间进行逻辑分析，而不是诉诸直观。这使得计算机能够按照逻辑推理，即在没有真正了解具体状况的情况下创建数学推理。我们在后面还会了解由约翰·罗杰斯·希尔勒（John Rogers Searle）设计的思维试验——“中文房间”。这个思维试验探索了机器翻译的算法，以推翻强人工智能（机能主义）提出的过强主张，试图说明遵循规则不能显示智力和理解力水平。

遵循数学游戏的规则，你就可以得到数学的定理。但是，这种进行数学证明的冲动是从何而来的？稍加试验就会发现，每个数字都可以写

成质数相乘的乘积，而且分解这个数的方法似乎有且只有一种。例如， $105=3\times5\times7$ 。3、5、7都是质数，除此之外没有其他质数的组合相乘可以得到105。在验证的过程中，更多的例子可以增强你对这个发现的信心，你会希望它总是有效。事实上，经过一段时间的验证以后，你可能认为证据是压倒性的，甚至可以把它作为一个公理。

但是，如果有一个非常非常大的数字有两种质数组合可以分解它呢？只是在这之前，你需要不断地验证，直到遇到那个非常非常大的数字。我认为这是标志数学不同于其他任何一门科学的最本质的点。若非如此，一位科学家想要说服其他科学家就不得不疲于奔命，不停地搜集数据，并依靠这些证据。但证明的存在意味着我们可以以逻辑推理的形式证明不会有任一例外的数字会打破这个理论。数学证明会告诉你为什么有且只有一种方法可以把数字分解为质数的乘积，这个证明将允许后来者直接将其作为公理来利用。

相较于同时期的其他人，古巴比伦人对数字和几何有更科学的运算方法。他们会对数字分解成质数乘积这样的运算结果感到满意，但他们会觉得没有必要拿出一个无懈可击的论证来解释为什么这一定是永真的。直到古希腊人发明了一种新的体系，把数学作为一门独立的学科，才让我们得以建立真理。

那么，这种证明的冲动是从哪里来的？它很有可能是社会演变的副产品，从权力集中的古埃及、古巴比伦，再到民主的古希腊。在古希腊人的日常生活中，讨论民主制度、法律制度，进行政治争论是市民生活的一部分。正是在古希腊，我们看到作家开始用逻辑论证来挑战权威和约定俗成的观点。

在这一时期出现的故事中，人类不再乐于被奥林匹斯山的诸神摆布，开始对诸神的统治提出异议和挑战。苏格拉底这样说：“未经审视的生活不值得度过。”他一直致力于论证真理与被接受的意见之间的区别。索福克勒斯<sup>[3]</sup>（Sophocles）笔下的安提戈涅向她舅舅的暴政发起了挑战。阿里斯托芬<sup>[4]</sup>（Aristophanes）在他的喜剧中讽刺了政治家的绝对权力。

这种对权威的挑战，向民主和以法律制度为基础的社会转变，需要人们拥有逻辑论证的技能。城邦的发展给了公民在社会中发挥作用的机会，这就使人们需要拥有新的技能来参与辩论。那时，诡辩家会到各个城市给人们上修辞课。亚里士多德将修辞学定义为“在任何情况下能

抓住说服对手机会的能力”。他阐明了一个公民需要具备什么样的素质，包括运用逻辑论证的能力和根据现有事实说服群众的技巧。

社会的这种变革激发了人们提出巧妙的数学证明形式。逻辑给了人说服别人的力量，这使用逻辑论证来表达自己的说服别人与数学证明同时发端。逻辑推理是如此强大，它使我们足以获得关于数字和几何的永恒真理：你可以证明每个数字都可以被唯一的一组质数分解；你可以证明质数是无穷多的；你可以证明在圆中，以圆的直径为边所得的所有三角形都是直角三角形。

很多时候，你会对这些真理的存在有一种预感，它们基于对数学的运用。例如，把奇数按顺序相加，得到的结果总会是某个数的平方，如 $1+3=4$ ， $1+3+5=9$ ， $1+3+5+7=16$ 。但这是一个真命题吗？古希腊人并不满足于发现奇数和平方数之间这种有趣的联系，他们想用逻辑推理的新工具来证明这是一个真命题。这就是利用逻辑推理来揭示数字运作的基本公理。

自此，人们开始了伟大的数学之旅。《几何原本》为后世2000多年的数学家提出的证明奠定了基础，这些证明解释了数字和几何的奇妙而独特的运行方式。费马证明了当 $N$ 为正整数且 $p$ 为质数并大于 $N$ 时， $N$ 的 $p$ 次方除以 $p$ 所得余数是 $N$ 。欧拉证明了 $e^{\pi i} + 1 = 0$ ——著名的“欧拉公式”<sup>[5]</sup>。高斯证明了每个正整数都可以分解为3个三角数（他在发现的旁边写下“Eureka”）。此外，我的同事安德鲁·怀尔斯证明了费马大定理。这些突破正是数学家工作的成果。数学家不是计算器，而是证明的构建者。

因此，本书的核心问题是：为什么计算机不能成为费马、高斯和怀尔斯一样的存在。对计算机而言，在计算方面显然没有任何人可出其右，但它构建证明的能力又如何呢？证明可以转化为一系列的符号，并为的一组符号与另一组符号之间的关系设置一个规则集。正如希尔伯特所讲的，你不需要知道符号的含义就能构建数学证明。这不正是让计算机参与证明的一个完美的设想吗？

数学家选择一个已有的数学命题，采用一种可被执行的逻辑进行运算，这时就会产生一个由新的符号序列组成的新的数学命题。这个命题可能已经在数学证明的列表中出现过了，因为我们可能通过其他的路径得到了它。但即使如此，对于数学家或计算机来说，这仍然是一种从已发现的定理中寻求新定理的有效方法。这不正是数学所追求的目标吗？数学不仅仅是不停歇的计算。如果按下计算机的“开始”按钮，它就不停



地输出通过运算、逻辑推理所得到的逻辑结果，那么它会不会让数学家集体下岗呢？

在创造新的事物方面，创造力发挥着不可替代的作用。自上而下的编程模式，将驱使计算机发现新的数学定理。对于计算机来说，关键在于所造物的价值，但这种价值从何而来？价值的导向和判断全都从人类创造和使用的数学思维中来。一个计算机的算法怎样知晓什么样的数学发现可以刺激你产生多巴胺和肾上腺素，从而让你感到兴奋呢？

对我这样的数学家来说，机器学习出现的自下而上的新编程模式，在让我感到兴奋的同时也让我深深感受到了潜在的危机：哈萨比斯和他的同事们正在开发的算法，可以从过往的人类数学经验中学习如何区分令人激动的定理和无聊的定理，而这反过来又可能会引导机器产生一个新的价值定理。这个定理可能会让数学界震惊，就像AlphaGo在棋类游戏界产生的震撼效果一样。

[1] 据加拿大数学家约翰·查尔斯·菲尔兹（John Charles Fields）要求设立的国际性数学奖项，于1936年首次颁发，常被视为数学界的诺贝尔奖（诺贝尔奖本身未设数学奖）。菲尔兹奖每4年颁奖一次，在由国际数学联盟（IMU）主办的四年一度的国际数学家大会（ICM）上举行颁奖仪式，每次颁给2~4名有卓越贡献的年轻数学家。获奖者必须在该年元旦前未满40岁，每人将得到15 000加拿大元的奖金和金质奖章一枚。——译者注

[2] 1862—1943，德国著名数学家，被称为“数学界的无冕之王”，是天才中的天才。希尔伯特领导的数学学派是19世纪末至20世纪初数学界的一面旗帜。——译者注

[3] 雅典三大悲剧作家之一。——译者注

[4] 古希腊早期喜剧代表作家，生于雅典。他熟悉希腊文学和艺术，与同时代的哲学家、文学家交游甚广。他对后世喜剧影响甚大，被称为“喜剧之父”。——译者注

[5] 欧拉公式是数学里最令人着迷的一个公式，它将数学里几个最重要的数字联系到了一起：两个超越数，自然对数的底 $e$ ，圆周率 $\pi$ ；两个单位，虚数单位 $i$ 和自然数的单位1；被称为人类伟大发现之一的0。数学家们评价它是“上帝创造的公式”。——译者注

## 第10章 数学家的望远镜

弗里德里希·尼采（Friedrich Nietzsche）

我们的写作工具参与了我们思想的形成过程。

尽管我有点担心计算机会让我丢掉工作，但我不得不承认，作为一种工具，它是一个“无价之宝”。当我们需要将一系列方程合并成一个方程时，手工计算是很难保证不出错的。但对于计算机来说，它就很擅长处理这种重复而机械且计算量庞大的任务。你只需要定义一套规则，剩下的就由计算机接手了。而且，在速度与准确性等方面，计算机是远超过手工计算的。正因为如此，近年来计算机的作用越来越重要，其应用领域也越来越广泛。

数学与计算机程序的算法紧密相关。因此，近半个世纪计算机常用于证明一些复杂的数学问题。20世纪70年代，计算机对“四色定理”的证明轰动了全世界。四色定理指的是“任何一张地图只用四种颜色就能使具有共同边界的国家着上不同的颜色。”也就是说，在不引起混淆的情况下，一张地图至少需要四种颜色来标记。

尽管此前很多人认为五种颜色就是下限，但计算机的发明大大加快了对四色定理证明的进程。1976年，数学家凯尼斯·阿佩尔（Kenneth Appel）和沃尔夫冈·哈肯（Wolfgang Haken）在前人的基础上用计算机证明了四色定理。阿佩尔与哈肯把地图的无限种可能情况简化为1936种构型，但是要靠人工逐一验证如此之多的构型是不现实的，所以才需要借助计算机进行验证。计算机根据程序指令逐一浏览地图并检查其是否满足四色定理。当时的计算机运算速度还不够高，整个证明过程的耗时超过了1000小时。

计算机只能执行指令，并无自主创造力。但是，想要证实程序中是否存在错误是很困难的。我们能在多大程度上相信计算机，这个问题一直困扰着人工智能领域的学者。当我们进入由算法主导的未来时，确保代码中没有未被检测出的错误，将成为一项艰巨的挑战。

2006年匹兹堡大学的托马斯·黑尔斯（Thomas Hales）教授在《数学年鉴》上发表了关于借助计算机证明著名的数学问题——“开普勒猜



想”的论文。简单来说，开普勒猜想就是对在空间中如何最密集地堆积圆球的解答。出于有效利用空间以及避免压坏水果的考虑，水果店店主一般会将水果层层交替堆叠，任意平面上的每个水果都与六个水果相邻，构成正六边形。像阿佩尔与哈肯一样，黑尔斯采用的也是借助计算机对足够多的案例进行穷举证明的方法。事实上，早在1998年，黑尔斯就曾宣布他的证明完成，并向《数学年鉴》评审组提交了论文、程序代码及相关资料，但该项证明的审核验证经历了漫长的时间。这是因为人类大脑的物理局限性，审核人必须得充分相信计算机的能力，就好比我們第一次乘坐飞机一样，心中难免惴惴不安。用了8年时间，数学家们证明了黑尔斯是正确的，但其确定性是99%。

对于数学纯化论者来说，这1%也是不可容忍的。这就好比，要证明你是牛顿的亲戚，可是家族谱系图里却缺少了关键的一环……人们质疑计算机证明数学问题的能力，并不是因为害怕计算机在未来会使得他们丢掉工作（早些年计算机只会按人类的指令执行操作，并不具备自主学习能力），主要是因为无法确定计算机程序是否存在潜在缺陷。我们该如何去相信计算机的证明呢？

数学家们就曾被程序代码中的缺陷困扰。1992年，牛津物理学家利用弦理论中的启发法对高维几何空间中可识别的代数结构数量进行了预测。对于该预测，数学家们持怀疑态度，因为他们觉得物理学不具备解释抽象结构的能力。当有证据表明这个猜想是错误的时候，他们觉得自己的怀疑是有道理的。然而，后来事实证明，否定这个预测的错误证据正是由一个有缺陷的计算机程序生成的。所以，错的是数学家，而不是物理学家——程序的错误把他们引入了歧途。几年之后，数学家们开始努力地证明物理学家的预测是对的（这一次数学家们把计算机排除在外了）。

这样的故事加剧了数学家们的担忧，他们担心计算机可能会让我们在结构不健全的“程序地基”之上建造精巧的“数学大厦”。但坦白讲，许多问题的证明往往都存在不足或错误，人类犯错的可能性通常比计算机更大。包括我本人发表的一些证明，后来也被发现存在一些漏洞。错误可以被修正，但遗憾的是，在证明的验证和审核阶段它们并没有被找出来。

证明的验证和审核非常重要，它是发现缺陷和漏洞的重要环节。这就是为什么数学界“千禧年大奖问题”的证明要经过两年的审核验证期——大家认为24个月的时间足够让错误暴露出来。以安德鲁·怀尔斯证

明“费马大定理”为例，在其证明方法付梓之前，审验人员发现了一个小缺陷。但怀尔斯和理查德·泰勒（Richard Taylor，曾是怀尔斯的学生）奇迹般地修正了这一缺陷。即便如此，在错误证明的基础上构建数学体系的情况也是屡见不鲜的。

许多新的证明极其复杂，以至于数学家们很担心一些潜在的错误难以被发现。以有限单群分类定理（classification of finite simple groups）为例，单群在有限群论中的地位，与素数在数论中的地位、原子在化学中的地位一样，它们都是构建各自所在世界的“砖块”。对于任意的有限群，我们可以将其分解为一系列单群，且分解方法是唯一的。通过研究这些“砖块”，我们可以进一步发现由它们所组成的物质的结构和性质。与当年化学家寻找新元素一样，数学家也开始了对于单群的寻找——列出一个单群的“元素周期表”，并证明这个“周期表”中包含了所有的单群。其中，“魔群”是最大的“散在单群”。“魔群定理”的证明散落在100多篇论文中，合计超过10 000页，涉及数百名数学家。单群的“元素周期表”中含有26个散在单群，对于是否存在第27个散在单群，人们总是持怀疑态度。对于这种类型的复杂证明，进行人力审核几乎是不可能的，那么，是否可以通过计算机程序来检验数学定理的证明呢？

新的问题又出现了，用计算机程序去检验计算机证明的步骤，是否可信？怎么确保计算机程序中没有缺陷？再用另一台计算机去查证吗？这会陷入一个永无休止的死循环。你怎么能确定你的方法正在引导你走向真正的知识的“圣杯”？真理的产生取决于你的证明方法。

正如哲学家大卫·休谟（David Hume）指出的，大多数科学研究都建立在归纳法之上——通过观察特定的例子来推断出一个普遍的规律或原则。为什么归纳法是一种产生科学真理的好方法呢？这主要是因为在归纳法里我们可以举出许多例子来说明。基于归纳法，曾产生了许多著名的科学理论，这反过来证实了归纳法确实是一种科学研究的好方法。

## Coq证明助手

在过去，数学问题的证明和验证过程全凭人工完成。而现在，越来越多的证明开始借力于计算机，但因为验证的过程既烦冗又复杂，并且工作量巨大，人类大脑的局限性决定了无法采用人工验证的方式判断其对错。因此，我们迫切需要一种解决方案，即通过构建新的程序来验证计算机证明的正确性。

20世纪80年代末，法国数学家皮埃尔·于埃（Pierre Huet）和蒂埃里·科昆德（Thierry Coquand）开始从事结构微积分（calculus of constructions）项目。该项目简称CoC，但很快又被称为Coq（法语里意为“公鸡”）。这个改动一方面是为方便记忆，因为在法国一直有以动物命名开发工具的习惯；另一方面是因为Coq是其开发者之一科昆德姓氏的前三个字母。Coq为验证数学证明而生，很快也成了验证计算机证明的重要程序，备受青睐。

2000年，微软研究院首席研究员乔治·贡蒂尔（Georges Gonthier）及其同事使用Coq对阿佩尔与哈肯的四色定理的计算机证明进行了验证，因为这是史上第一个需要计算机才能完成的证明（假定Coq不存在任何缺陷）。然后，他们也使用Coq去验证了阿佩尔和哈肯自己所写的证明部分。

人类手工证明与计算机证明不同，手工证明过程中会跳过一些烦琐或众人皆知的步骤，而计算机却依赖于明确、细化的步骤才能正确执行指令。这类似于写小说和写保姆指导手册的区别。前者不需要对主人公的每一个动作都解释得一清二楚，而后者则需要尽可能地明确和详尽，包括一天中婴儿的食谱，以及吃饭、睡觉、上厕所的每一个细节。

计算机用了5年的时间进一步自动识别并验证人类证明的过程。这期间，人们惊讶地发现了在第一次证明中被忽略的数学知识。

Coq与原始的计算机证明相比，更应该信任谁呢？当然是前者。越来越多的计算机证明被Coq所验证，使我们更加确信Coq是可靠的。这就像我们通过归纳法验证数学中的基本公理一样。这就像任取两个数A和B，如果 $A+B$ 都等于 $B+A$ ，那么 $A+B=B+A$ 就是正确的。用一个计算机程序来验证多个计算机证明，比编制一个特定的证明程序或者进行人工证明更值得我们信任。

贡蒂尔团队验证完四色定理后，紧接着开始了对奇阶定理（odd order theorem）的验证工作。奇阶定理是对称性研究最重要的指导定理之一，通常被认为是有限单群分类的基石。像化学里的元素周期表一样，有限单群是构成数学有限群论“元素周期表”中的基本元素，所有的对象都由有限单群构成。具有素数边的正多边形（如正三角形、正五边形）是该周期表中的元素。此外，该周期表中还有一些复杂且独特的对称元素，如旋转了60次的正二十面体、需要196 883维线性空间才能表达的“魔群”等。“魔群”具有的元素个数超过了构成地球的原子个数。

该定理指出，任何奇阶对称结构的基本组成单元都是素数多边形，此外再无其他结构。如果把对称物体分为奇阶和偶阶两种，那么该定理就等于涵盖了其中的一半，意义重大。

奇阶定理的原始论文有255页，占据了《太平洋数学期刊》的全部篇幅。在它出版之前，大多数证明最多只有几页，一天内即可掌握。这个冗长复杂的证明，对每一位数学家来说都是一个挑战。因此，其中是否存在细微的缺陷或错误，始终无法考证。

Coq对复杂数学定理的证明过程，一方面可以检验Coq的能力，另一方面能帮我们树立足够的信心。但将人工证明转换成可验证的计算机代码这一过程并不容易。

贡蒂尔略带腼腆地回忆道：

第一次开会讨论时，我向团队里其他成员宣布了我的宏伟计划，他们流露出不可思议的表情，就像是我得了妄想症。奇阶定理的证明过程太过复杂，验证它最初被认为是不可能的。做这个项目的真正原因，是为了充分理解数学理论的构建过程并使之与Coq充分融合。

会议结束后，团队里的一名程序员查看了原始证明，随后向贡蒂尔发来一封邮件：“17万行代码，1.5万个变量，4300个函数。好玩，太棒了！”微软剑桥研究院团队用了6年的时间完成了证明。当项目即将结束时，贡蒂尔兴奋地说，经过无数个不眠之夜，他终于可以放松一下了。

贡蒂尔说：“数学是最伟大的浪漫主义学科之一，即便是天才，也得掌握所有知识才能激发灵感，理解一切。”但是，人类的大脑存在物理上的局限性。他希望他们所做的一切能够叩开人类与机器彼此信任、持续合作的新时代“大门”。

## 人脑的极限

年轻的数学家们开始意识到，数学研究变得更为艰难了：学科分支越发密集，问题越发复杂。攻读博士学位的3年时间，只够去理解导师所给题目的含义。随后，再花费数年时间去研究、探索，运气不错的话，会得到一些研究成果。然而，你发表的论文却面临着没人能审核它。

审核别人发表的论文是得不到太多报酬的，但期刊论文的审核必须经过同行的评审。职称评定也以公开发表在《数学年鉴》或《l'IHES数学期刊》这类文献中的论文积分为基准。因此，有一个像Coq证明助手这样的系统就非常重要了。

一些数学家认为我们目前正处在一个新旧时代的交替期——数学的发展虽然受到人类大脑局限性的制约，但借助于计算机，我们对数学的探索已远远超出了人脑的思维范畴。

伟大的数学家们能够用他们睿智的头脑，借助于纸和笔这些极其简单的工具，构造出像“魔群”这样具有196 883维的对称体，这是人类的奇迹。但数学家们终将会老去，就像中世纪的泥瓦匠，其精湛的技艺将伴随身体的死亡而从人世间消失。如果很难找到通往“新奇迹”的方向，人们终将失去创造的原动力。

费马大定理的证明长达数百页，跨越3个世纪，这说明人类有足够的耐心。当你努力去证明一个极其复杂的猜想时，隐约会有一种突破人类大脑物理极限的感觉。数学是无限的，而人的能力是有限的。但即便如此，我们常会为自己所做的努力感到吃惊，因为我们用数学的方式证明了“数学海洋的广阔无边”。

有一个问题几乎困扰了我15年之久。每次推演时，总是在即将得到解决方案的关键时刻，我的大脑容量就不够用了，它给我“即将宕机！”的警告。距离成功仅一步之遥，却难以取得突破。就像现象与本质之间隔着一张“渔网”，它制约着我们，让我们难以冲出迷雾，得到光明。当几代数学家致力于黎曼假设的证明而不得其解时，人们开始怀疑，是否这样的证明已超越人脑的极限。

著名数学家哈代多年来一直试图证明黎曼假设，后来他自嘲

道：“每个傻瓜都能提出有关质数的问题，而最聪明的人却无法解答。”奥地利数学家、逻辑学家库尔特·哥德尔（Kurt Gödel）有过论证：数学中包含了许多没有经过证明的真理。能否用新的公理去证明那些未被证明的真理呢？哥德尔早在1951年就发出了警示，他认为我们可能会越来越难以掌控现代数学的发展方向：

人们创造出了一套庞杂且仍在扩展的公理系统，但人们研究它的目的越来越说不清楚……的确，在现代数学中，这些更高层次的理论成果实际上无法投入使用，这有可能与它们无法证明某些基本定理有关，例如黎曼假设。

鉴于我们可能即将触及人类自身能力的极限，一些数学家已意识到，如果希望人类文明持续进步，我们将需要更多的机器辅助。就好比登上珠穆朗玛峰之巅，我们可能只需要一个氧气罐，但如果人机不能结合，我们永远无法登上月球。

以色列数学家多伦·泽尔伯格（Doron Zeilberger）认为：数学家只用铅笔和纸张工作的日子即将结束。20世纪80年代以来，他一直使用计算机撰写论文。他将自己的由AT&T（美国电话电报公司）生产的计算机命名为“Shalosh B.Ekhad”（希伯来语中3B1的意思），并坚持将这位机器伙伴作为论文的联合作者。泽尔伯格认为，人们之所以不愿倚重人机合作的方式，是因为“狭隘的人本主义”在作祟，这种偏执与其他形式的偏执一样，阻碍了人类发展的脚步。

大多数数学家认为他们孜孜以求的目标非常深奥，是计算机难以企及的。换言之，他们不仅希望能得到真理，并且希望探求真理背后更多的内涵。如果计算机在无法真正理解数学的情况下就能验证数学真理，他们会觉得非常荒谬。

获得菲尔兹奖的数学家迈克尔·阿蒂亚（Michael Atiyah）曾说过：“我们的理想是探究数学真谛，而不是利用机械执行指令的计算机推演论证。”另一位菲尔兹奖获得者泽尔曼诺夫（Zelmanov）也表示赞同：“只有所有数学家都认可的证明方法才是真正有效的，所以我对机器证明方法的前景并不看好。”当然，我们也不会认可只有一位数学家采用的证明方法。泽尔曼诺夫说的有道理吗？如果数学证明方法只有生成它的机器能够理解，我们真的可以相信吗？

起初，多伦·泽尔伯格对这种观念也非常理解，但最终对其不屑一



顾。他承认，让他乐此不疲的是在整个数学证明过程中得到所有的证据。对他来说，这就是生活，而生活是错综复杂的。他相信如果人的头脑可以找到一个证据，那么它一定是显而易见的：

二三十年后，人类可以通过计算机轻松完成大多数事情。在数学领域里，利用计算机完成很多工作已经变成现实。现在，人类很多刚发表的论文就已过时了，其实完全可以用算法来替代人类完成这些工作。现如今我们遇到的很多问题已经变得毫无意义，但是我们还是继续在做，仅仅因为这是人类可以做的事情。

对于数学领域的现状来说，这是相当令人沮丧的评估。但这是真的吗？我当然觉得有些论文进入期刊是因为我们需要出版物，但这并不总是坏事。为了做某事而做某事所带来的意想不到的成果已经多次证明，无目标驱动的研究有时是收集真正的新见解的最佳方式。

像许多业界同人一样，乔丹·艾伦伯格（Jordan Ellenberg）认为人类未来在数学领域中仍然会发挥至关重要的作用：

我们非常擅长搞定计算机无法做到的事情。想象一下未来，如果目前所知道的一切定理都可以通过计算机得到证明，那我们就可以去探索计算机无法解决的其他问题，这有可能成为未来的“数学”。

但人类的许多研究成果不是向前而是横向平行延伸的。在某些领域我们确实达到了临界点，想要超越珠穆朗玛峰的高度就必须借助一台机器。这对保守派来说是一种震撼的观念冲击（也可能包括我自己）。他们不愿承认的是，人类再也不可能仅使用笔和纸来探求数学的奥义了。

## 沃沃斯基的愿景

弗拉基米尔·沃沃斯基（Vladimir Voevodsky）是明星数学家，他仅用传统的纸笔工具就在数学研究领域成绩斐然，但后来他开始倡导数学家使用计算机技术辅助研究工作。我和他是在牛津大学相见的，当时我们正在尝试请他为我们工作。人们都认为他是获得菲尔兹奖的不二人选，于是牛津大学决定尽早给他一个待遇优厚的工作职位。沃沃斯基在自己的工作研讨会上提出了一个数学研究的新理念，他似乎想引导大家使用一种能够证明那些困扰几代数学家问题的全新数学语言。他认为这种新的数学语言不是某种学术研究思路的延续，也不是几种观点的简单组合。

我之前在书中谈到了三种创造力：探索型创造力、组合型创造力和变革型创造力。通过引入一种全新的视角来改变一个领域的面貌，沃沃斯基的创造力确实具有变革性。聆听他的观点，你不禁会问：“这是从哪里来的想法？”

原来，激发他这种非凡创造力的竟是一些意想不到的东西。交谈期间，当我知道他在选择未来工作地点的一个重要考虑因素是能否获得他需要的药物时，我感到非常震惊。我所说的药物并不是很多数学家的首选药——咖啡因。（正如著名的匈牙利数学家保罗·厄多斯所说：“数学家是将咖啡转化为定理的机器。”）他要求我们必须源源不断地给他供应一些相当“过瘾”的B类药物作为他去牛津大学的交换条件。

我从未真正觉得药物会对提升逻辑思维的缜密程度大有裨益，但沃沃斯基认为安非他命<sup>[1]</sup>能使他产生大量幻觉，可以使他激发出极大的灵感。当我看到咖啡因和安非他命对蜘蛛织网产生影响时，我开始有些相信他的方法。正常蜘蛛可以迅速织出形状规则的网，而摄入咖啡因的蜘蛛织的网就像一团乱麻。沃沃斯基后来获得了菲尔兹奖，并接受了普林斯顿大学高级研究所的职位，但他早期的成功方式却引发了一场生存危机。

他说：“我意识到下一个数学问题的证明可能在学术领域不会产生太大的影响了，因为数学即将面临一场危机，或者更确切地说，似乎是两个危机。”

这两个危机的其中之一涉及“理论数学”和“应用数学”的分离。随着研究预算越来越紧张，政府不得不面对把科研经费投向何处的两难困境。一些政客开始质疑：为什么要把大量资金投给那些对社会没有实用价值的研究工作。因此，沃沃斯基极其重视向外界阐述他所从事的研究工作将来会对社会产生的深远影响。

另一个危机更像是一场生存危机，它与数学研究变得日益复杂有关。即使数学家们能够精通各自研究的细分领域，但对他人的研究过程和成果缺乏足够的了解。因此，数学家正在变得越来越孤立。早在1739年，大卫·休谟（David Hume）就在他的《人性论》中阐述过社会背景也能对学术研究产生重要影响：

在他的研究领域中，没有一个代数家或数学家能像他那样专注。他对自己发现的任何真理都充满信心，哪怕它仅仅是一种可能的存在。每当他推演出自己的证明成果，尤其在得到更多同行的赞许和学术界的普遍赞誉与认可时，他的信心就会更加坚定。

沃沃斯基还认为，学术期刊的论文迟早会由于过于复杂而无法得到详细的论证，而这将导致期刊文献中出现未检测到的错误。从某种意义上说，数学是一门深度学科（具有复杂的层阶结构），一篇论文通常将之前许多篇论文的结论作为依据，这种错误的叠加效应将会得到非常危险的结果。

在意识到这两个潜在的危机后，沃沃斯基决定放弃为他赢得名声和荣誉的研究领域，转而致力于研究数学如何避免发生潜在灾难的问题。他的第一个挑战是运用数学方法来解决其他领域的问题。他从小就对生物学很感兴趣，所以他想知道他开发的数学工具是否可以为公认的非数学领域提供新的见解。他花了几年时间，尝试通过分析种群的当前基因结构来追本溯源，但他试图破解生物学之谜的研究项目最终搁浅。与其所擅长的数学领域大相径庭的是，他发现自己并没有深入探究生物学问题的工具和技能。

他说：“到2009年我才意识到我的研究成果竟然毫无用途。到目前为止，它也许是我学术生涯中最大的遗憾。虽然我为该项目做了大量的工作，但最终还是失败了。”

经过多次深入灵魂的反思，他转向了他所发觉的另一个危机：理论数学正在变得日益复杂。沃沃斯基认为，如果人类无法检验彼此的证明

成果，那么我们可能需要寻求机器的帮助。那些批判他观点的人对迫在眉睫的危机不为所动。对于许多数学家来说，谈论使用计算机辅助研究似乎就是误入歧途。大多数数学家继续坚信，在自身敏锐直觉的引导下，人类思维才是促进方程和几何研究发展并可获得解决方法的核心力量。

当沃沃斯基四处寻找合适的工具时，他发现唯一可行的计算机系统是法国的Coq系统，它能够引导证明。最初他无法理解它是如何运作的，所以他向高级研究院申请由他教授Coq系统的课程。“如果你想要深入了解某些知识和概念，就试着备课并教会别人”。渐渐地，他明白了，计算机学者所使用的程序设计语言，起初看似难懂，实则是他早些年作为数学家时运用过的一种抽象世界的范式。

他似乎找到了同时化解两个危机的方法。首先，他发现看似呆板机械的数学范式竟然完美地构建了计算机世界；其次，他可以利用这种“新语言”来创建一套全新的研究数学领域的基础方法，计算机将在其中发挥核心作用。

绝大多数数学家认为沃沃斯基对未来数学的设想太具颠覆性，甚至有些人认为他已“堕入深渊”。那些仍用纸笔计算的学者（可能现在也使用计算机来检验常规计算结论）和那些想用计算机来证明新定理的学者之间仍然存在较大分歧。虽然使用计算机检查证明结论的方法正在被更多人接受，但数学研究的方向和方法依然掌握在人类手中。也就是说，计算机还不具备在数学研究领域的独创能力。

然而，沃沃斯基认为：“我不知道持有这些观念的人前途何在。这些保守迂腐的观念终将被抛弃。新的数学研究方法首先会被一小部分人运用，然后逐渐普及，最终成为标准，紧接着就会出现在数学专业研究生的课堂上，进而推广到本科教育阶段。这个过程大约需要几十年，接下来会发生什么就不得而知了。”

沃沃斯基还将这种人机互动比作玩计算机游戏：“你告诉计算机尝试这个，它尝试了，然后把执行结果反馈给你。有时结果会出人意料，这非常有趣。”

遗憾的是，沃沃斯基没有机会看到这些改变了，2017年他死于动脉瘤，享年51岁。

总之，在沃沃斯基的启发下，我也开始对计算机能够激发数学创造力持乐观态度。鉴于数学和音乐之间存在某种密切的联系，我想通过研究AI是如何创作音乐来思考计算机怎样能对数学研究产生更大的作用。毕竟，正如巴赫的学生洛伦茨·米兹勒·冯·科洛夫（Lorenz Mizler von Kolof）所说的，“音乐就像会发出声音的数学”。

[1] 精神类药物名称。——译者注

## 第11章 音乐：声响的数学之旅

戈特弗里德·威廉·莱布尼茨（Gottfried Wilhelm Leibniz）

音乐——人类心灵在不自知的运算中所体验到的快乐。

1964年，菲利普·格拉斯（Philip Glass）在巴黎跟随娜迪亚·布朗热<sup>[1]</sup>（Nadia Boulanger）学习作曲时，每一节课都要从巴赫开始。《赋格的艺术》<sup>[2]</sup>是他们课程的一个关键部分，每周的专业课上格拉斯都要学习一部巴赫的作品。和声是多声部音乐的音高组织形态，是音乐的基本表现手段之一，是与对位<sup>[3]</sup>相对应的技术范畴。四部和声是多声部主调音乐最基本和最常用的声部组合形式，四部和声的“四部”与人声的四种声乐类型是一致的。在布朗热的指导下，格拉斯重新构建了这些四部和声，并使这些新加入的音和谐地与全曲融合起来。布朗热认为，所有伟大的作曲家都必须从学习巴赫的音乐开始。

我心中总是有一个小小的愿望，成为一个作曲家而不是一个数学家。在我的数学生涯里，音乐一直是我的好伙伴。例如，巴赫的作品以严格的对位著称，他十分中意对称的结构；巴托克的作品很多都以黄金比例为结构基础，他非常喜欢并善于使用斐波纳契数列。当我尝试探索数学新领域的时候，实际上我的大脑是在寻求一种新的模式和结构。无论是对称还是黄金比例，这些都是让作为数学家的我感到兴奋的东西，而这些作曲家同样被这些模式和结构吸引。这可能就是巴赫或巴托克的音乐有助于我思考的原因。有时，作曲家是本能地或者不自知地被数学的模式和结构所吸引，而他们并没有意识到这些数学模式的意义；有时，他们主动去寻找新的数学思想来作为他们作品的框架。

在和作曲家艾米丽·霍华德（Emily Howard）讨论几何时，我有了一个想法：作为交换，我给她讲解双曲几何<sup>[4]</sup>，她给我上作曲课。她欣然接受了这个建议。不久后，我在咖啡馆上了我的第一堂作曲课。

就像一个新人写手面对一张完全空白的稿纸一样，面对空空如也的五线谱我感到了恐慌。艾米丽平静地解释道：“每一位作曲家在开始创作一部作品时，都需要先构思一个框架，以便进行创作。”她建议我们从中世纪的复调音乐开始。在中世纪，有一种复调作曲技法叫作有量卡农<sup>[5]</sup>（prolation canon）。这种作曲技法是从一个简单的节奏开始，单



声部进行；随后，第二声部以一半的速度、同样的节奏加入；再然后，第三声部以两倍的速度、同样的节奏加入。尽管主题一致，通过这样的作曲技法就可以让你听到三个声部有着不同的节奏。当了解这种作曲技法后，去听这样的作品时，你立马会在脑海里识别到这种模式并将这三个声部组织起来。

我的家庭作业是：创作一段简单的节奏，用有量卡农的作曲技法写一首弦乐三重奏。对我来说这是一个简单的任务，我可以很容易地写出这首曲子结构对应的数学方程式： $X+2X+1/2X$ 。在我创作的过程中，我强烈地感觉我像是一个园丁：从无到有创造出来的一小段节奏就像一颗种子，我将它安置在五线谱上。通过应用艾米丽教给我的算法，这个种子在我手中不断地变化——它在生长，而算法帮助我生成新的部分来填充其余的声部。新生成的部分与原始的种子有很强的联系，但它不只是复制原始的种子，或是它的一部分。看到我的音乐花园在这个简单的规则下变得欣欣向荣，我感到非常满意。

正是这次作曲的经历，让我理解了算法和作曲之间的密切联系。算法是一组规则，它可以接受各种各样的输入，通过将规则应用于输入的内容，可以得到结果。初始输入就是种子，算法是种子生长的方法。我们在前文已得知，运用欧几里得算法<sup>[6]</sup>可以求得两个正整数的最大公约数。算法可以有不同的图像，通过分析这些图像，你可以知晓这些算法的内容。有一些算法可以生成分形图形：从一个简单的几何图像开始，反复将一个数学方程应用到图像上，最后就会出现一个复杂的图形。

前面我应用在音乐创作上的算法也具有类似的特性。格拉斯早期的一部作品或许表明了，算法是这位作曲家工具箱里的一个关键工具。他的作品《1+1》是为一人演奏所作，演奏者需要在桌面上敲击出一系列的节奏，所发出的声音会由接触式的麦克风放大。这个作品的主题动机（音符组合的模式）有两个：A是两短一长的节奏，B是一个长的节奏。在格拉斯的指导下，演奏者可以选择某一种算法（规则）来组合这两个主题。

这样，演奏者就被赋予了自己选择算法的自由。格拉斯还做了一些不同算法的范例，例如ABAABBBAAABBBBBB.....这个算法的规则是每当主题A增加一次，主题B就相应增加两次。我想会有很多人这样批评格拉斯：“得了吧！这哪里是音乐？太过单调乏味了吧！”但对我而言，这个作品向我展现了音乐的核心——当你听到这首作品时，你的大

脑会意识到：作品的内容不是随机的，也不是简单的重复。尝试逆向开发作品并从中找到核心模式是一种乐趣。我相信正是这样，使音乐世界和数学世界的联系变得如此紧密。

因此，作曲家在创作的时候要进行艺术或科学的双重考量：发明新的算法，用以创作有趣的音乐；选择不同的种子（即主题动机）输入到算法当中。鉴于现实工作中算法正在不断地提升音乐创作的质量，这可能是计算机成为作曲家的关键吗？

[1] 1887.9.16—1979.10.22，法国著名女音乐教育家、作曲家、指挥家。  
——译者注

[2] 被看作巴赫（1685—1750）在晚年悉心于音乐艺术最深层探索的思想和实践的最后总结，是其一部未完成的杰作。整部作品由14首赋格和6首卡农组成，它们是由同一个主题发展而来的。这部作品的主要目的是要在一部音乐作品里，尽可能地深度发掘复调音乐对位法的所有可能。——译者注

[3] 即通常所说的“复调”。——译者注

[4] 也称罗巴切夫斯基几何、波利亚-罗巴切夫斯基几何、罗氏几何，是一种独立于欧几里得几何的几何公理系统。——译者注

[5] 又称mensuration canon或proportional canon，是一种复调作曲技法，指卡农的不同声部以相同或相似的节奏（主题动机）、不同比例的速度进行。——译者注

[6] 又称辗转相除法，用于计算两个正整数的最大公约数，应用于数学和计算机两个领域。欧几里得算法和扩展欧几里得算法可使用多种编程语言实现。——译者注

## 巴赫：首位音乐程序员

布朗热坚持将学习巴赫的作品作为学习音乐创作的起点，她坚持的原因之一是，算法在巴赫创作音乐的过程中发挥了显而易见的重要作用。在某种程度上，我觉得巴赫应该是首位音乐程序员，因为巴赫的许多作品可以用数学的语言描绘出来，但其蕴含的算法要比中世纪的复调音乐复杂得多。虽说《音乐的奉献》源于腓特烈大帝要求巴赫即兴创作，但这部作品清楚地证明了上述观点。

尽管普鲁士国王腓特烈大帝以他在军事上的功绩而闻名，但他一生都非常热爱音乐。他的父亲在他还是孩子的时候，曾试图摧毁他对音乐的爱好和追求，但并无效果。登上王位后，腓特烈大帝经常在波茨坦的宫殿里为彰显他的文治武功，招揽最优秀的音乐家举办盛大的音乐会。在这些音乐家里，有巴赫的儿子卡尔·菲利普·埃马努埃尔·巴赫（Carl Philipp Emanuel Bach，即C.P.E.巴赫），他是乐团的羽管键琴首席演奏家。

1747年，62岁高龄的巴赫去看他儿子时创作了《音乐的奉献》这部作品。当时的交通不便，在经历了好几天辛苦的旅程后，他终于到了儿子家，这时他已经累瘫了。腓特烈大帝听说巴赫已经到达，立刻下令让巴赫入宫，邀请他参加宫廷音乐会。据说腓特烈大帝十分青睐西尔伯曼钢琴，他把当时仅有的15架全部收入囊中，摆放在宫殿的各处。

收到宫廷的传召，为了不让腓特烈大帝久等，巴赫立马动身前往，甚至没有换下旅行的行装。他一到宫殿，腓特烈大帝就带着他参观那些收藏的钢琴。巴赫惊人的即兴创作能力广为人知，因此腓特烈大帝在宫廷音乐会上给了巴赫一个音乐主题，让他用其收藏的钢琴根据该主题创作一首乐曲。

腓特烈大帝给巴赫的这个主题，并不是一个普通的主题。这是一个缺乏音乐性的主题，几个音符之间的音程关系毫无关联，它们被生硬地堆砌在一起。一般人是无法用这种主题作曲的。20世纪的作曲家阿诺尔德·勋伯格（Arnold Schoenberg）对《音乐的奉献》的巧妙构造感到惊讶，他这样评价：“这不仅仅是单一对卡农模仿。”换句话说，用任何经典的对位规则都无法解释这部作品。此故事有另一个版本：腓特烈大帝和巴赫的儿子编造了这个困难的无旋律性的主题，来挑战巴赫。C.P.E.

巴赫是巴赫（约翰·塞巴斯蒂安·巴赫，即J.S.巴赫）的次子，因为巴赫在音乐界的影响甚大，他厌恶一直以来都活在父亲的光环之下。他认为父亲的作品是老派的，他想创造一种新的音乐风格。所以，这次即兴作曲也许是暴露他父亲作曲风格和方法中缺点的机会，好让他出丑。也许正如勋伯格所说，他们希望“享受巴赫掉进这个精心设计的陷阱里的无助感”。如果真是这样的话，那就真的是适得其反，搬起石头砸了自己的脚——巴赫坐下来，用这个怪异的主题，即兴创作了令人惊叹的三部赋格曲。

赋格曲是一种更复杂的卡农或轮唱，很多人在学校里应该唱过这种歌。卡农的唱法是这样的：一半的同学先开始唱，几拍过后，另一半同学再开始唱同一首歌；一个声部的曲调自始至终追逐着另一个声部，直到最后的一个小节；最后的一个和弦，歌声融合在一起，给人以一个神圣的意境。《伦敦大火》（London's Burning）和《雅克兄弟》<sup>[1]</sup>（Frère Jacques）都是很好的例子。

这里所使用的算法非常简单，具有非常明显的几何性质。首先，创作旋律导句是卡农的基础，你需要把它写在五线谱上。算法是一个规则，这个算法应用在旋律导句上生成一首和谐的曲子。该算法的工作原理是，先获取旋律导句的副本，然后在时间轴上向右移动，重复相同的乐句（见图11-1）。这种算法选择模仿导句进入的时间不同会产生不同的效果。这有点像罐子上的条形图案，可以复制、移动和重复。就像罐子上的图案，轮唱曲也可以在第一声部、第二声部之后再添加第三声部。



图 11-1

如果把卡农的算法写成一个数学公式，那么我们设旋律的导句为X，然后设延时为S，那么演奏或演唱这首卡农的数学表达式就是 $X+SX+SSX$ 。该算法创建了一个单一主题的三声部卡农。

赋格在卡农的基础上进一步发展了，成为多声部多主题动机的多声复调作品。巴赫在作品中喜欢使用的一个手法是：不仅在二声部的时间上做延迟，还对二声部的主题做音高移动的处理。他还会在作品中运用对称规则：二声部可能是将主题旋律反过来使用，就像镜面反射一样。结合所有这些规则，人们可以构建一种算法，来应对如腓特烈大帝故意为难巴赫这样的难题：以一个怪异的主题为基础创作出一首和谐而复杂的乐曲。巴赫已经掌握了这种算法，该算法可以帮助他解决现场即兴创作赋格曲的难题，来应对腓特烈大帝的挑衅。

腓特烈大帝对这种即兴表演印象深刻，但他想知道巴赫是否能在乐曲上继续添加声部，创作六声部的赋格曲。这是前所未有的。尽管难度很大，巴赫却不想不战而降。创作一部六声部的赋格曲需要更多的时间和精力去构思，不仅仅是坐在键盘前即兴创作就能完成的。所以他回到莱比锡，重新对腓特烈大帝的主题进行变奏创作，以两首赋格曲、四乐章三重奏鸣曲和十首卡农构成了整部套曲，在两个月后呈送给了腓特烈大帝。这其中的赋格曲采用的就是现在为人所称道的、令人惊叹的“里切卡尔”（ricercar）。

巴赫创作的这部套曲里的十首卡农都是以腓特烈大帝所给的主题为基础的，构成每一首乐曲的都是基础的主题动机和一个数学规则或算法。每一首乐曲都像是一道数学题，演奏者必须解开这道数学题才能进行演奏。例如，巴赫在一首单行乐曲的末尾加了一个倒置的谱号。这个上下颠倒的谱号是巴赫这首乐曲算法的关键，也是演奏者将这个算法应用的关键。算法的内容是：将谱子上下颠倒过来，按照谱面所记演奏出一条旋律，然后再将这条旋律和原有的旋律同时奏出，这样就得到了一首二声部的钢琴曲。该算法是应用于原始旋律主题的规则，目的是为原始声部添加额外的声部。正如图像识别的算法适用于任何照片一样，只要有一个基础的主题动机旋律，这个音乐算法也可应用其上。

《音乐的奉献》开篇的十首卡农分别用了不同的算法，这些算法在数学的层面上对原始的主题进行了发展。这十首卡农是为后面的里切卡尔做热身，而里切卡尔向我们展示了巴赫如何完美地以一个简单的主题为基础，运用简单的数学算法创作出一部精妙绝伦、异常复杂的作品。这里面应用到的技法有：正格卡农的同度卡农和不同度卡农，变格卡农的增时卡农、减时卡农、倒影卡农，等等。巴赫将这些令人眼花缭乱的规则巧妙地结合在一起，创作出了六声部赋格曲。我们的大脑一边紧张地识别乐曲进行的模式，一边分析这个模式的运行规律，但未来是未知



的，所以在分析的同时我们对乐曲接下来的进行模式做出预测是异常困难的。正是这种知与未知点燃着我们的兴奋点。正如哈里森·伯特威斯尔（Harrison Birtwistle）所言：最后一个音符没有演奏完，音乐就还没有结束。

巴赫当时是否意识到他正在进行数学游戏？在我来看，很明显他知道自己在做什么。因为其中涉及太多有关于数学的部分，我们很难以偶然或是潜意识中的无意间使用来解释。巴赫晚年时申请进入他的学生米兹勒创办的米兹勒音乐学社。该社致力于研究和探索科学与音乐之间的联系，曾公开发表过题为《作曲学习中数学的必要性》的论文。作为会员的巴赫，当然也是支持、认同数学和音乐之间具有紧密的相关性的。

C.P.E.巴赫对父亲的赋格相当不屑，他声称自己“对枯燥的数学一点都爱不起来”。为了证明赋格除了运用一些小伎俩以外没什么其他有意义的技巧，他甚至设计了一个作曲小游戏——“音乐白丁也可做出二声部复调6小节的乐曲”。玩家手中有两页乐谱，每页有看似随机的一系列音符：第一页为右手准备，即高音声部；第二页为左手准备，即低音声部。玩家所要做的就是随机选择一个音符作为开始，然后是这个音符后的第9个音符，再然后是第18个音符、第27个音符……直到将所有的音符都使用过为止。C.P.E.巴赫设计的关键点在于首音的选择，无论从哪个音符开始，只要遵循继续选择其后第9个音符，这样简单的规则就可以让任何音乐白丁构建出一段可以被接受的对位复调作品。这是多么完美的机器算法！

《音乐的奉献》经常会被搬上音乐会的舞台，但是我从未听过哪场音乐会演奏了C.P.E.巴赫这样“创作”出的作品，这表明成功的音乐作品可能不仅仅是机械地遵循一套规则就可以。

莫扎特被认为其有一套类似于C.P.E.巴赫的算法来谱写华尔兹舞曲：一种叫作“骰子音乐”（*musikalisches würfelspiel*或*musical dice game*）的游戏算法，使用一组骰子生成16小节的华尔兹舞曲。讲解该游戏算法的《骰子音乐》于1792年首次出版，这时莫扎特已经去世了一年。有人怀疑这个游戏算法根本就是从出版商尼古拉斯·西姆洛克（Nikolaus Simrock）设计的，他只是盗用莫扎特的名字以提高销量。

这款游戏共176小节，按照11×16的布局排列。第一列是开始的第1小节，有11种不同内容的小节可选择。游戏方法是，掷两个骰子，在得到的点数基础上减去1，这样我们就能得到1~11的点数了。举个例子，



若掷出两个6点，那么就选中第一列的第11小节。第二列代表整首曲子的第2小节，同样用掷骰子的方法选择。用这种方法继续，直到完成所有列的选择，最终拼合成16小节的华尔兹。

令人震惊的是，使用这个系统可以生成1116种不同的华尔兹，即约4600万亿<sup>[2]</sup>首华尔兹。一首一首听，一直不停，你需要两亿年才能听完。早期的“算法艺术家”会使用这样的技巧——将预先设定的结构和随机性元素结合。莫扎特的作曲游戏精妙地将176小节（固定结构）和掷骰子（随机性）结合起来，无论怎样都能拼合成令人信服的华尔兹圆舞曲。当然，并非所有的变化都是悦耳的，有些组合显然比其他组合的效果更好。令人遗憾的是，莫扎特并没有给出评价拼合作品的标准。对我而言，这是开放式算法的问题之一。

[1] 中国版是《两只老虎》。——译者注

[2] 45 949 729 863 572 161。——译者注

## 艾米：人工智能作曲家

我喜欢做这样一件事：当听到收音机播放音乐时，我会在公布是谁的作品之前猜是谁的作品。一天早上，我坐在书桌前听到一首乐曲，我很快意识到这是巴赫的风格，这首曲子应该是巴赫的作品。曲终时，播音员说这首曲子是由一个算法创作的，这令我十分讶异。让我震惊的不是这首作品让我以为它是巴赫创作的，而是在短短一首乐曲的时间内我被所听到的打动了。一段代码真的能做到吗？让我好奇的是，其背后的算法是如何让我认为所听到的是伟大的巴赫创作的乐曲。

巴赫的作品是大多数作曲家开始学习创作的起点，也是大多数计算机开始学习作曲的起点。那天我在广播中听到的那首曲子，是由一位一直在苦苦寻找灵感的作曲家按照简单的代码规则创作出来的。大卫·柯普（David Cope）正在创作一部委约歌剧作品，可是他灵感枯竭，连一个音符都写不出来。当记起阿达·洛夫莱斯曾经预言的“分析机可能会生成任意复杂程度、精细程度的科学的音乐作品”时，他决定在乐思枯竭的状况下求助于算法，探索阿达的预言。

那是在20世纪80年代初，计算机还不像现在我们见到的这样。他将打孔卡片输入IBM计算机，而音符就是输出。后来他回忆道：“早期实验的结果确实不尽如人意。”但他坚持了下来，并前往斯坦福大学学习了计算机音乐的相关课程。随着委约作品交稿日期的临近，他决定将他在计算机上做的试验付诸现实。

如果有一种算法可以准确地理解、把握他的创作风格，那么每当他的创作陷入困境并无法继续的时候，这个算法就能给他提供与他创作风格相兼容的建议。即使算法给出的建议是荒谬的，至少算法会帮助他理解并找到可能会是更好的选择。这个算法将成为激发创造力的催化剂。柯普为他的这一套试验命名为“音乐智能试验”（experiments in musical intelligence），简称EMI。从算法试验中诞生的“作曲家”被命名为艾米（Emmy），起这个名字一方面是避免与英国百代唱片公司（EMI）重名，另一方面是柯普想让艾米更人性化。

柯普在创作歌剧上下了7年的功夫，在艾米的帮助下，他用两周时间就完成了这部歌剧的创作。歌剧《摇篮坠落》（Cradle Falling）发表时，他决定不让外界知道这部作品是在艾米的帮助下创作的，以免使乐

评人产生偏见。发表两年后的1987年，歌剧首次公演时，柯普很高兴地发现这部作品获得了他职业生涯中最好的一些评论。一位乐评人说：“这作品太感人了！无疑是一部现代的经典名作！”这样的反馈鼓舞着柯普继续与艾米合作创作作品。

如果该算法能够学习柯普的创作风格，那么它是否可以经由训练“变成”那些“经典大师级的作曲家”呢？举个例子：算法可以在学习巴赫或巴托克的创作风格后，创作出那些风格的作品吗？柯普相信每一段音乐都内含了编码和指令，用以创造出其他相似但微妙不同的音乐片段。这里所面临的问题是如何将这些指令具体化成代码。

在艾米的帮助下，他开始为每位作曲家建立一个与他们的风格相对应的数据库，比如他们具有专属特性的音乐语言语汇和技法。音符就是字母，但是什么样的单词可以与某位作曲家的专属语言特征相对应呢？柯普分析的关键概念之一是标志性主题动机，这是一个由4~12个音符组成的序列，它会反复出现在同一作曲家的多部作品中。例如，在莫扎特的钢琴协奏曲中，我们会发现有一种模式反复出现，它被称为“阿尔贝蒂低音”<sup>[1]</sup>。它通常出现在音乐的二声部，由三个音按13231323的顺序组成（见图11-2）。



图 11-2

这个模式就会汇入到与莫扎特风格相对应的数据库中。在柯普分析的所有作曲家中，莫扎特的作品尤其富有特色。这些特征可能会以不同的速度、节奏和音高出现，但对于发现潜在的模式，数学是很拿手的。这有点像无论你以怎样的方式将一个球抛向空中，这个球始终按照抛物线方程描述的路径运动。

柯普的分析揭示了作曲家的作品带有强烈的富有个人特色的模式性。从巴赫到莫扎特，从肖邦到勃拉姆斯，从格什温到斯科特·乔普林<sup>[2]</sup>，每个人似乎都有自己偏爱的特定的动机。也许这并不奇怪。为什么在广播中听了几个小节后，即使以前从未听过这首曲子，我还是经常能

猜出作曲家是谁呢？就像盲人品酒师一样，我在研究一些关键的指标。而在音乐中，这些指标就是音符的模式，它们就像画家标志性的笔触一样。有些作曲家，如巴赫，甚至在乐谱上用音符来签名。巴赫在其最后的巨作——《赋格的艺术》中最后一首未完成的赋格里，以“音乐签名”的方式把自己的姓名“BACH”留于其中。在德国乐理体系里，与B对应的唱名是降si，与A对应的唱名是la，与C对应的唱名是do，与H对应的唱名是si。

在将作品分割成一个个单元和标志性动机，即形成每个作曲家的数据库之后，柯普的算法转向了他所说的“重构”。识别、拆解、分析一个复杂结构是一回事，找到一种方法将经过拆解的构件重新组合构造成一个全新的结构是另一回事。柯普本可以选择使用像莫扎特的《骰子音乐》游戏那样的随机过程，但是随机组合不太可能反映出作曲家在创作中所想展现的情感张力，所以他在程序中又增加了一个步骤：他为每一部分都创建了热图。

作曲家经常把各种元素组合在一起，他们称之为乐句或乐节。柯普尝试用一种他称为“SPEAC”的方法来将这些模式抽象化。如果数据库是字典，那么“SPEAC”就是作曲家使用字典中的单词编写乐句的方式。“SPEAC”确定了乐句的五个基本组成要素：

**Statement（声明）：**“简单存在”的乐句，只能做重复。

**Preparation（准备）：**以出现在S或其他要素前为前提，修改其含义的要素。

**Extension（扩展）：**一种扩充、延展S的方法。

**Antecedent（先导）：**有重大暗示、引导作用并要解决问题的乐句。

**Consequent（解决）：**将A未解决的问题解决的乐句。C通常与S具有相同的和弦或旋律片段，但是它们具有不同的含义。

许多古典主义时期的作曲家都会使用这种套路来作曲，有时他们是在不知不觉的状况下使用，但通常他们在学习作曲的时候就学会了这种作曲技法：这个和弦需要在曲子的这个位置得到解决；接下来的和弦会让人觉得回到了主音，有一种安定感；这个和弦会进一步提高亟待解决

的程度，让人觉得非要解决不可。柯普可以在SPEAC的帮助下分析一首作品，因为每个作曲家都有自己独特的套路。例如，图11-3是柯普对斯克里亚宾的一首钢琴曲的分析：



图 11-3

柯普的这一基本程序建立伊始，他就用其测量音程的和谐度。若纯八度音程<sup>[3]</sup>和纯五度音程是和谐的，不会造成很大的紧张感，这在数学中也得到了证实——两音之间的振动频率比是小整数比：纯八度是1:2；纯五度是2:3。钢琴上两个相邻键之间的音程（小二度或半音）就会造成很强的紧张感。同样地，在数学中也反映了这一点：它们的振动频率比是更大的数字之间的比（15:16）。通常在一段音乐中听到了这种高紧张感的音程后，你会知道它们接下来必然朝着低紧张度的方向解决。

柯普将这些规则输入到系统中，用以帮助艾米从给定的作曲家作曲风格的大型数据库中构建新的乐曲。艾米的重组规则是将片段按照一定的指导原则组合在一起，这些片段必须与柯普的SPEAC分析编码相匹配。例如，如果片段B与片段A具有相同的起始动机，但是片段B的发展指向了一个新的方向，那么组合的方式就是片段A+片段B，而不是片段B+片段A。

当许多不同的片段适合组合时，就需要做出选择了。柯普不喜欢使用随机性去选择，而更喜欢使用数学公式去选择。数学公式会提供一个“任意结构”<sup>[4]</sup>来控制所做出的选择，就像“无法解释的可预测性”指导“绘画傻瓜”作画一样。1993年，柯普和艾米准备发行他们的第一张专辑《设计的巴赫》（Bach By Design），这张专辑收录了艾米创作的巴赫风格的乐曲。专辑中的曲目因为难度太高，人类演奏员难以胜任，所以他们不得不求助于一台既能作曲又能演奏作品的电脑。可惜，这张专

辑并没有受到乐评人的好评。

“当我看到这些评论时，我感到非常沮丧，因为他们的关注点在于作品的演奏技巧，而不在于作品本身。”考虑到作品本身并没有受到攻讦，他觉得自己还是有勇气把这个项目继续下去的，于是他在1997年推出了第二张专辑。这张专辑里收录的作品的风格包括他分析过的其他作曲家，如贝多芬、肖邦、乔普林、莫扎特、拉赫曼尼诺夫和斯特拉文斯基等。但不一样的是，这次专辑中的曲目全是由人类音乐家演奏的。这一回，乐评人的反应要积极得多。

[1] 阿尔贝蒂1710年生于威尼斯，1740年卒于罗马。曾师从A.比菲和A.洛蒂，学习歌唱和对位法。在音乐史册上，他的奏鸣曲非常有名。这些奏鸣曲的左手部分惯用分解和弦音型作为伴奏（其奏出的次序为低音、高音、中音和高音）。这种分解和弦音型，后人称为“阿尔贝蒂低音”。——译者注

[2] 1868—1917，美国作曲家和钢琴家，是历史上不多见的黑人音乐家，被誉为“拉格泰姆（ragtime）之王”。Ragtime是一种黑人旋律结合切分音法（syncopation）循环主题与变奏乐句等法则的早期爵士乐，盛行于第一次世界大战前。其发源在圣路易斯和新奥尔良，而后在美国的南方和中西部也流行起来。它影响了新奥尔良传统爵士乐独奏与即兴演奏的风格。这种爵士乐不但在黑人乐手与乐迷间流行，也被美国白人中产阶级所接受。——译者注

[3] 音程指两个音级在音高上的相互关系，即两个音在音高上的距离，其单位是“度”。——译者注

[4] 在构造“结果树”的过程中，来自于“源树”的元素可以被任意地过滤、重组，或被添加任意结构。——译者注



## 模拟游戏：音乐图灵测试

柯普算法输出的结果能通过音乐的图灵测试吗？这些作品可以以假乱真吗？为了进行测试，柯普与道格拉斯·霍夫斯塔特<sup>[1]</sup>（Douglas Hofstadter）合作，在俄勒冈大学举办了一场音乐会。霍夫斯塔特是一位计算机科学家，著有经典著作《哥德尔、艾舍尔、巴赫：集异璧之大成》<sup>[2]</sup>（Gödel, Escher, Bach: an Eternal Golden Braid）。在这场音乐会上，演奏了三首曲目：第一首是巴赫所作但并不广为人知的作品，第二首是艾米以巴赫风格创作的作品，第三首是音乐理论教授史蒂夫·拉尔森（Steve Larson）以巴赫风格创作的作品。这三首曲目由拉尔森的妻子，钢琴家威妮弗雷德·科纳（Winifred Kerner）按随机顺序演奏。

对于拉尔森的作品，观众们的评价是：该作品是由一台毫无情感的电子计算机所创作的。这令拉尔森很沮丧。然而，他的失望很快就被另外一个结果所取代了：算法艾米所做的作品被大家投票表决为“巴赫本人的作品”，而巴赫本人的作品居然被大家投票表决为“拙劣的伪作品”。

“我发现我彻底被艾米搞糊涂了。”霍夫斯塔特若有所思地说道，他试图弄明白其中的道理，“唯一令我感到安慰的是，我明确地知晓艾米没有自己的创作风格，它的创作风格是由模仿对象决定的。但这还是令人有些不舒服。有爵士乐手这样说，音乐在很大程度上是由‘即兴重复乐段’组成的。如果大部分情况都是这样，那么对我打击最大的是，音乐的世界比我想象的要狭隘太多了！”

柯普继续在世界各地进行着这样的试验，而参与者的反应开始令他身心俱疲。在德国，一位音乐学家被激怒了，其在音乐会结束后威胁柯普，并控诉他“杀死”了音乐。这位音乐学家的块头很大，比柯普要重90多斤，柯普觉得自己只有躲在周围的人群当中才是安全的。而另一场音乐会的情况是，一位教授在演出结束后找到柯普，激动地说道：“这是长久以来我听到的最优美的曲子之一。”在音乐会后的讲座中，这位教授才意识到这个音乐作品都是由电脑算法创作的。这一信息彻底改变了教授对该作品的印象。讲座结束后，他再次找到了柯普，并斩钉截铁地表示“这一切简直是太肤浅了”。此外，他改口道：“从开始的第一秒，我就知道这是电脑合成的！在这样的东西里我感觉不到任何的勇气、情感、灵魂！”柯普被他的出尔反尔、反复无常惊呆了。同一部作品，却

让他的态度发生180度大转变，而其中唯一的改变是，他是否知道作品是由计算机代码生成的。

还有一次，霍夫斯塔特演奏了两首曲子，一首是肖邦的，另一首是艾米创作的肖邦风格的曲子。观众中有许多作曲家和音乐理论家，但他们都认为电脑生成的曲子是肖邦所作。其中一人事后写信表达了钦佩之情，并描述了自己投票后知晓结果时的情形：“大家都倒吸了一口凉气.....当时，我们只剩下了惊喜和惊诧。我从来没有见过这么多作曲家和理论家一下子从沾沾自喜中清醒过来（包括我自己）！这简直太美妙了！”

霍夫斯塔特对艾米创作的肖邦风格的作品讶然道：“这的确是艾米新创作的，但确实具有肖邦的风格，它在情感的表达上也并不空洞和乏味。这太令我震惊了。一个从未活过一分钟、从未聆听过任何音符、从未有过任何情感的程序怎么能创作出如此饱含深情的音乐作品呢？”

柯普认为他的算法之所以能如此成功，是因为该算法触及到了人们创作音乐的核心。他说：“据我所知，没有哪首富有表现力的音乐是不使用算法的规则来创作的。”尽管这种说法会使人们感到困惑，甚至激怒人们，但许多作曲家都会同意这种观点。只有那些真正的门外汉才不敢承认他们的情绪状态可以被代码左右。柯普曾这样透露：“在我看来，原创力（不是重组或形式推演的结果）来自人类的灵魂或与上帝的某种神秘联系，这种说法完全是荒谬的。”

这可能是对的，但我认为重要的是人们要认识到，尽管音乐可能比我们通常理解的更加数学化和编码化，但这并不会影响、剥夺音乐情感的本质。当我谈到数学和音乐之间的联系时，人们会非常不安，认为我正在把他们喜欢的音乐变成一种纯粹客观、冷酷的东西。但是，他们并没有真正理解我的意思。在我来看，与其说音乐像数学，不如说数学像音乐。我们所喜爱的数学包含着巨大的情感内容，这让我们被其深深吸引。那些有能力欣赏数学语言的人会随着一个曲折的证明而心潮澎湃，就像我们很多人会被一段动人心弦、饱含深情的音乐所激励、感动一般。

我认为，在我们大脑中运行的“人类代码”，已经进化到对构成自然界的混乱的抽象结构具有高度的敏感性。当我们听音乐或探索具有创造性的数学问题时，我们被暴露在最纯粹的结构形式中，我们的身体会在情感上做出反应，用以表明对这种结构的认识。是什么帮助我们区分随

机音符序列和我们所承认的音乐呢？根据信息论之父克劳德·香农的研究，我们的部分反应可以归因于这样一个事实：非随机序列在底层基础上可以用某种算法压缩数据，而随机序列并不能。音乐与噪声的区别在于其内含的算法不同。什么算法决定某种声音让人觉得是值得听的音乐呢？许多人认为，音乐在某种程度上是对生活经历的情感反应。而算法都是在隔音的录音棚里完成作品的，没有与周围的世界产生联系、互动。没有体验，就不要指望它能真的模仿做出大师之作。霍夫斯塔特肯定相信，或者可能希望情况是这样的：

一个能像肖邦或巴赫那样创作音乐的“程序”必须要独自环游世界，在生活的迷宫中摸索前行，感受生活的每一刻。它必须理解寒冷的夜风带来的欢乐和孤寂，拥有对“执子之手，与子偕老”的渴望，眺望那些永远无法接近的遥远城镇，并拥有关于死亡、心碎的痛感和重生的愿望。音乐意义之源就在其中，也只在其中。

但是，其实是听众把他们自己的情感经历带入到他们所听到的音乐中。听众、观众或读者，在一件艺术作品的创作过程中所起的作用常常被低估。许多作曲家认为这种情感反应来自于音乐的结构。但是，你无法为情绪编程。菲利普·格拉斯认为情绪是随着作曲的进程而自发产生的：“我发现音乐几乎总是带有一些情感色彩，但这似乎并不受我意图的支配。”

音乐和情感之间的关系一直是作曲家们“魔力”的源泉。斯特拉文斯基的作品极富表现力，他的见解在这个问题上尤其有说服力。他认为情感不属于音乐，而属于听众：

音乐本质上是无法表达任何东西的，无论是一种感觉，一种心态，一种心理情绪，一种自然现象.....如果真的如此这般，而你认为音乐似乎表达了一些东西，那么这只是一种幻觉，而不是现实。这只是一种附加属性，是我们经过长期的默许或根深蒂固的思维定式作为标签和惯例，给音乐增添、强加了这种属性。简而言之，我们已经无意识或习惯性地混淆了音乐的本质。

那么，为什么音乐似乎违背了如此强大的情感反应？也许作曲家已经成功地识别了大脑编码某些情绪的方式。这些编码情绪的频率或音符对于不同的人来说可能是不同的。大多数人都会同意，我们所谓的小调式与悲伤有关，这是后天的反应还是与生俱来的反应？作曲家可能会选择一个小调来捕捉一种情绪，这意味着一种直接的编码，但音乐理论还

没有发展到我们能充分理解这种编码如何运作的阶段。所以，作曲家很可能是在“暗箱”中进行创作的。正如斯特拉文斯基和格拉斯所说的，他们创造了结构，情感从结构中浮现出来。

许多作曲家喜欢建立规则或结构来帮助他们产生音乐思想：巴赫喜欢写谜题一样的赋格曲；勋伯格开创了十二音体系；巴托克使用斐波纳契数列作曲；梅西安在《时间结束四重奏》中用质数变值作为四重奏的框架；菲利普·格拉斯师从布朗热学习作曲期间感觉学业异常艰难，但他最终脱颖而出，开创了简约主义音乐。

斯特拉文斯基认为约束是他产生创造力的关键：

我的自由是把有限的精力分配到我力所能及的每件事中去。正因为如此，我将会走得更远：我越是限制自己的领域，越是给自己设定障碍，我的自由越将变得伟大和有意义。

我的作曲老师教给我了一套可以帮助我作曲的规则，让我开始了自己的小小音乐之旅。在使用了有量卡农之后，我又为自己创建了一些规则，并完成了一个算法来指导我作曲。我曾经看到过约翰·凯奇的轶事，他经常作完曲后，在作品首演的时候才第一次听自己的作品。我对此种方式很好奇，也想这样听听我对数学的重新想象会是什么样子。

但当坐在钢琴旁听到自己创作的弦乐三重奏时，我很失望。我所遵循的规则本打算让曲子有一个有趣的逻辑来吸引听众听下去，但实际上曲子听起来怪怪的。我真的不知道这是为什么。当然，认为音乐和数学一样有正确或错误的唯一答案这种想法是愚蠢的。在对最初的结果感到失望之余，我开始违反我设定的规则，打乱了我写下的音符，以期创作出一些听起来更具乐感的东西。我无法解释为什么我要做出这些改变，我把自己交给了更深层次的东西——身体与音乐，潜意识和人性。

这是一个重要的教训。作曲是规则、模式、算法以及其他诸多因素的融合，这就是霍夫斯塔特所说的“环游世界，行万里路”所得到的东西。正是这种神秘的东西开始渗入我写下的音符，开始赋予它生命和美感。

这些结构是否需要通过情感的感知来了解？如果需要，计算机怎样感知情感？如果音乐是在编码情感，那么这些代码能用于模拟计算机的情感状态吗？也许，创建艾米的2万行代码已经成了其中的一部分。

霍夫斯塔特对艾米创作的肖邦风格的作品有情感上的反应，这是对2万行代码的情感反应吗？难道这段代码不是像肖邦创作的作品那样捕捉情感的吗？

把艾米输出的音乐称为人工智能创作的音乐是个骗局。艾米的创作依赖于作曲家数据库。作为作曲家，柯普拥有职业的敏感性和专业的分析工具，可用于选出与作曲家风格相对应的元素，并重组这些元素。艾米的大部分创意来自柯普和历史上伟大的音乐大师的作品。

柯普使用自上而下的编码过程构建了艾米，是柯普编写了所有的代码用来输出音乐。我们现在处于这样一个阶段：我们可以将更具自适应性的新算法应用到作曲家的原始数据里，而无须经过人类音乐分析的烦琐过程，并且我们可以训练这些算法从零开始学习音乐理论。那么，机器学习的算法能够从头开始并创造出与历史上的伟大作品相媲美的作品吗？答案就像音乐学中经常说的那句话：让我们再次回到巴赫。

[1] 又译为侯世达，1961年诺贝尔物理学奖得主罗伯特的儿子。他是美国著名学者、计算机科学家、印第安纳大学计算机科学和认知学教授、观念与认知研究中心主持人，也是哲学、心理学、比较文学、科学史与科学哲学副教授。——译者注

[2] 该书获得1980年普利策文学奖（此奖是美国出版界的最高奖项）。作者通过对哥德尔的数理逻辑、艾舍尔的版画和巴赫的音乐的综合阐述和对比，引人入胜地介绍了数理逻辑学、可计算理论、人工智能学、语言学、遗传学、音乐、绘画的理论等多方面内容。——译者注

## “深度巴赫”：从头开始再生作曲家

巴赫写了389首四部和声的圣咏合唱，格拉斯对其进行过重构，柯普对其做过分析。巴赫著名的《约翰受难曲》包括好几首圣咏合唱。如果你正在寻找巴赫对数学痴迷的例子，你会在他所做出的选择中找到，并发现巴赫痴迷于数字14。在巴赫的时代，许多欧洲思想家和哲学家对卡巴拉<sup>[1]</sup>（kabbalah）很感兴趣，它包括将字母转换成数字并探索单词之间的数字联系以推断更深层次的联系。巴赫很惊奇地发现，他姓氏的字母按照排序翻译成数字加起来是14（2+1+3+8）。这成了他的“数字”签名，就像足球运动员球衣上的号码。例如，巴赫要加入其学生米兹勒建立的米兹勒音乐学社，他其实可以早早入会，但他一直在等，等到第13个人入会后他再入会。他还发现了一些有趣的方法把数字引入他的作品。在《约翰受难曲》中，共有11首圣咏曲。我们来看一下前10首圣咏的小节数，是这样的：11、12、12、16、17、11、12、16、16、17。

接下来第11首圣咏是关键：有28（即 $2 \times 14$ ）个小节。将前10首两两配对，第1首和第10首，第2首和第9首，依此类推……我们得到了：11+17=28，12+16=28……如果以对称的方式相配对，小节数之和都会是28！这是巧合吗？我看不太可能！

创作这些合唱，巴赫通常会先安排路德教派众赞歌旋律在女高音声部，然后将其他声部进行组合，以协调旋律。柯普对其作品进行分析，然后将各声部的协调关系编入他的算法。他了解巴赫在和声中所使用的规则，但是，计算机只依靠本身能从原始的数据中学习和声的规则吗？做和声题的练习就像是玩一个异常复杂又非常需要耐心的游戏，或是做一个开放式的数独游戏。每一步，你都需要决定你的男高音声部下一步的走向。旋律走向高还是低？音域是否合适？节奏、速度怎样安排？在你考虑其他两个声部时，你需要考虑的就更多了，你必须要保持整体的协调，牵一发而动全身！

作曲专业的学生在做和声练习时，老师一定会强调很多规则。比如，一定要避免平行五度或八度<sup>[2]</sup>。平行五度将会削弱两个和弦的独立性，导致和声效果变弱，听觉上会感觉一下子从立体声变成了单声道。早在13世纪，在作曲的规范里就禁止使用平行五度，直至今日这还是作曲界的共识。



格拉斯回忆有一次上课，他的老师布朗热问道：“你最近身体怎么样？没生病吧？头疼吗？身体不舒服要去看医生，精神压力大了可以去看心理医生。我可以给你假，我们可以调课。”格拉斯答道：“我身体挺好的，没什么问题。”布朗热突然转过办公椅面向他，拿着他这周的和声作业，尖声咆哮道：“那这是怎么回事！你怎么解释！”果然，格拉斯在自己作业里的女低音声部和男低音声部里看到了隐伏五度<sup>[3]</sup>。

打破传统规则是创造性思维的标志。在AlphaGo的制胜一招中，我们能看到这一点。同样地，我们发现巴赫有时也会打破规则，使用平行五度。但这会不会让作品变糟呢？正如我的作曲老师艾米丽向我解释的，创作的乐趣之一就是打破规则，这是实现创意的最佳机会。

和声有一种二维的特性：和声必须在垂直方向上有意义，而旋律本身在水平方向上也必须有逻辑并且和谐。对人类作曲家来说，创作和声作品并将这两个维度结合起来是一个考验。

那么，由机器学习驱动的新算法可以解决这样的问题吗？巴赫创作技巧的秘密能否通过他的389首圣咏解码出来？测试这个问题的一种方法是做统计分析：根据这一个音符，猜测可能性最高的下一个音符。例如，在不同的作品里，ABCBA这样的旋律片段作为和声的一部分出现了好几次，那么你可以对A后面那个音进行统计分析。在作品BWV<sup>[4]</sup> 396中紧跟A的音符是#G，音高降低了半音。然而，在作品BWV 228里紧跟A的音符就跃进到了F。通过建立这样的统计分析，可以创建一个音乐骰子游戏，不同的权重代表不同的音符可能出现的概率。假设巴赫有8次选择了使用#G，有4次选择了F，那么算法就有2/3的概率选择#G。这有点像算法DeepMind学习如何玩《打砖块》游戏：算法该朝哪个方向移动球拍，移动多少能获得胜利。应用于音乐的算法只不过把球拍换成了或高或低的音符。

柯普在着手识别作曲家的标志性乐句时发现，这种方法的难点在于确定音符的数量：如果太少了，那么满篇都是；如果太多了，乐句就会被过度确定，以致输出变成了复制原作品。此外，除了音高，你还需要考虑节奏模式。

根据我们听音乐的方式，从左至右跟着时间轴，根据以前的节奏模式构建，这是最简便的办法。但这并不是对一件作品进行统计分析的唯一方法。在弗朗索瓦·帕切特（François Pachet）和弗兰克·尼尔森（Frank Nielsen）的指导下，音乐专业学生盖坦·哈德耶勒斯（Gaëtan

Hadjeres) 为他的博士论文开发了一种算法——“深度巴赫”(DeepBach)。该算法旨在透过现象看本质, 将巴赫的作品二维化, 并进行分析。这就像在拼图中移除一块, 并分析其周围的图像, 然后我们就能猜到巴赫往这空白处填了什么。因此, 它不是在时间轴上向前组合, 而是向后查看。这是解迷宫的一个典型技巧: 从结束开始, 试着如何回到起点。<sup>[5]</sup>当然我们也可以选择中间的部分, 问问“深度巴赫”怎么填补。

基于这种多维度的分析导致, “深度巴赫”创作出了比以前的算法所做的在结构上更加连贯的圣咏。那些算法漫无目的, 只是在时间轴上向前曲曲折折地组合。然而, “深度巴赫”的这种分析仍然是局部层面的分析: 其以每个音符为中心向四周扩散, 一组一组地观察并进行分析, 但是“音群”的大小是受限制的。在“深度巴赫”的算法中, “音群”的范围是以某音符为中心的前后各四拍。那么, 这个算法有多成功呢?

盖坦和他的导师将巴赫的圣咏分为两部分: 80%用来训练算法, 20%用作测试数据。志愿者被请来配合试验, 他们将听到“深度巴赫”和真正的巴赫的作品。志愿者必须在听完后判断该作品是计算机所作还是人类(巴赫)所作。志愿者的音乐教育背景被考虑到了, 因为这会严重影响评估的可靠性: 学过作曲的人的耳朵要比没经过训练的人的耳朵灵, 能听见他们听不到的声音。

总体结果令人震惊: “深度巴赫”的作品有50%被认为是巴赫的。学习过作曲的志愿者成绩稍好, 但仍然有45%的“深度巴赫”作品被认为是巴赫所作。这令人印象深刻。标准是无情的, 只要有一点点小纰漏就能被识别出来。巴赫在他的作品中没有犯任何错误, 然而他25%的作品被判定为机器生成的。这同样令人印象深刻。并不是我傲慢自大, 我觉得圣咏可能是巴赫作品中最乏味的。虽然这些圣咏是他反复推敲出来的, 但无法感动我。

任何试图向大师学习的项目的关键难点在于缺乏良好的数据。389首圣咏听起来很多, 但实际上只勉强够学习之用。在计算机图像识别的领域, 算法会有数百万张图像训练自己。而在“深度巴赫”这个算法中, 只有区区389个数据采集点, 而大多数作曲家远没有巴赫这样多产。测试巴赫的圣咏是很有实践意义的, 因为其提供了一组非常相似的单一现象的例子。但当你更广泛地观察作曲家的作品时, 你会发现这里有太多的变化, 以于一台机器无法从中吸取足够的有效信息, 甚至会迷失学习方向。也许, 这就是人类创造的艺术不受机器进步影响的终极安全锁

——好作品的数量太少，机器无法学习如何复制。当然，它们可以生成穆扎克音乐<sup>[6]</sup>，但无法创作高质量的音乐。

[1] “卡巴拉”是犹太教的神秘哲学，传说它最原始的根源来自埃及文化。生命之树，就是“卡巴拉”思想的核心，它被视为神创造宇宙的蓝图，也有人称之为神体的构造图。他们认为它不只是一个存在于纸上的图样，而且是一个真正存在的三度空间的宇宙，也就是我们存在于其中的这个宇宙。卡巴拉思想以生长于天国的“生命之树”来象征宇宙全体，其意义非常难解，真要说的话，生命之树意味着广大的宇宙、身为小宇宙的人体，以及达到神之境界的精神遍历。生命之树大概可以分为3个支柱、10个原质、4个世界、22条路径等基本结构。卡巴拉学者使用生命之树作为创世的示意图，从而将创世这个概念发展成为一个完全的现实模型。——译者注

[2] 平行五度（parallel fifths）指的是乐曲的两个声部隔开纯五度平行进行。八度亦同。——译者注

[3] 在分部写作中，两声部同向进入纯五度，叫作隐伏五度。这种进行中隐含着平行五度。八度亦同。——译者注

[4] BWV代表巴赫·韦尔科·韦尔泽奇尼斯（Bach-WerkeVerzeichnis），他对巴赫的所有作品进行了分类编号。——译者注

[5] 即反推法。——译者注

[6] 即安静柔和的背景音乐。这样的音乐在过去被称作“电梯音乐”，因为我们经常在电梯里听到它。这种音乐是专门设计出来让人放松心情或增加活力的。有时候，你甚至没有意识到正在播放这种音乐，但其总是会让你有所反应。近年来，我们在越来越多的地方都能听到它，因而它又被称为“衬托音乐”。——译者注

## 第12章 歌曲的创作公式

维克多·雨果（Victor Hugo）

音乐能言不能言，亦能言，不得不言。

我是一名小号演奏员，但是我从未掌握即兴爵士乐的技能。我在管弦乐队中按照乐谱演奏完全没有问题，但要成为一名爵士乐手，那就要求我同时成为一名作曲家。我曾经见过一位作曲家，他可以一边和乐团的其他同事正常交流，一边作曲。我一直非常佩服这样的天才。

在学习爵士音乐的过程中，我意识到实现好的即兴演奏有一个很大的令人困扰的问题。一般来说，一首爵士乐会有一组标准的和弦，这些和弦会随着乐曲的演奏出现并变化。小号手的任务是在变换和弦时吹奏一条与和弦相吻合的旋律线条。对于小号手来说，他吹奏的旋律必然是一个音符到另一个音符，同时还要保证能组成旋律，所以演奏爵士乐就像在二维迷宫中描绘一条线一样。和弦决定了垂直方向的动态，而他刚刚演奏的旋律决定了水平的动态。随着爵士乐即兴成分的增加，音乐变得更自由，同时和声的发展也会变得更加流动，这就要求小号手必须能敏锐地觉察并预判钢琴手的发展意图，其发展的方向是由前面和声的发展决定的。判断即兴演奏者水平的标准是，他是否会倾听并预判出主奏乐器的发展意图且配合得天衣无缝。

创建一台能够做到这一点的机器似乎并不是不可能，但是要克服很多问题，这是像艾米这样的算法作曲家没有遇到过的。即兴爵士乐算法必须在实时交互中同时完成播放与对新材料的处理和响应。

马克·列文（Mark Levine）所著的《爵士乐宝典》（The Jazz Theory Book）是许多青年音乐家初学时的经典用书。列文经常和20世纪最伟大的即兴爵士乐演奏家迪兹·吉莱斯皮（Dizzy Gillespie）、弗雷迪·哈伯德（Freddie Hubbard）一起演奏。诚如列文所说：“一首伟大的爵士独奏是由1%的魔力和99%的可解释、可分析、可分类、可操作的东西组成的。”这99%都是可以放进算法里的。

迈尔斯·戴维斯（Miles Davis）的Kind of Blue是我最喜欢的爵士乐专辑。那么，我们离创造“深蓝”（DeepBlue）还有多远呢？



## 普希金、诗歌和概率

弗朗索瓦·帕切特（François Pachet）在年轻时，梦想成为一名音乐家，这样他就可以像自己的偶像一样，能写出热门歌曲，还能抱着吉他帅气地自弹自唱。尽管在作曲方面没少下功夫，但最终他还是成了一个人工智能工程师。在巴黎担任索尼计算机科学实验室（Sony Computer Science Laboratory）主任期间，帕切特发现人工智能领域的学习工具可以帮他完成作曲的梦想。他使用概率论中的马尔可夫链公式写出了世界上第一首人工智能即兴爵士乐，并创造了世界上第一个可以进行爵士乐写作的算法。

迄今为止，我们在许多的算法中都可以看到马尔可夫链在起着重要的作用，这些算法是很多应用的基本工具：从模拟化学反应、经济趋势，到互联网导航，再到评估生物种群、人口动态。可是，俄国数学家安德烈·马尔可夫（Andrey Markov）选择检验自己理论的，不是科学，而是普希金的诗。

马尔可夫的发现源于其与另一位俄国数学家帕维尔·涅克拉索夫（Pavel Nekrasov）的争论。概率论的核心之一是“大数定律”<sup>[1]</sup>——如果你有一枚硬币，而每抛一次硬币都完全独立于上一次抛硬币，那么当你多次抛硬币时，正面和反面出现的次数之比就会越来越接近于一半对一半的比例。抛掷硬币4次，全是正面的概率是 $1/16$ ，但随着抛硬币次数的增加，偏离对半分概率的可能就会降低。

帕维尔·涅克拉索夫和马尔可夫的观点恰恰相反，他认为，现实世界中的事物是相互依存的（比如人的行为），所以现实中的事物并不恰好符合数学模式或分布。如果统计数据遵循大数定律，那么它的假设必是每个事件都是完全独立的。涅克拉索夫试图用这样的例子来证明，俄国的犯罪率统计是遵循大数定律的，但事实上，罪犯决定犯罪都是由个人意志决定的。

马尔可夫对涅克拉索夫错误的逻辑感到沮丧，他评价涅克拉索夫的论调是“对数学的滥用”，并决心证明其是错误的。因此，马尔可夫需要建立一个模型，在这个模型中，结果的概率取决于以前发生的事件，但长期来看仍然遵循大数定律。抛硬币的结果并不取决于以前抛硬币的结果，所以这不是马尔可夫理想的模型。但是，如果增加一点依赖关系，

使下一个事件取决于刚刚发生了什么，而不是整个系统如何影响了当前事件，又会怎么样呢？每个事件的概率仅取决于先前事件的一系列事件被称为马尔可夫链。预测天气就是一个例子：明天的天气肯定取决于今天的天气，但并不特别依赖于上周的天气。

我们来看以下模型。天气有晴天（S）、阴天（C）和雨天（R）。如果今天是晴天，那么明天就有60%的可能是晴天，30%的可能是阴天，10%的可能是雨天。但是如果今天是阴天，那么明天天气的可能性就会变化：明天下雨的可能性就变成了50%，阴天的可能性是30%，晴天的可能性是20%。在这个模型中，明天的天气只取决于今天的天气。即使我们已经有两周连续的晴天也没有关系，只要今天阴天，那么明天就有50%的概率下雨。这个模型的最后一部分是，如果今天是雨天，那么明天有40%的可能是晴天，10%的可能是阴天，50%的可能是继续下雨。让我们用矩阵来描述这些概率：

$$\begin{pmatrix} SS & SC & SR \\ CS & CC & CR \\ RS & RC & RR \end{pmatrix} = \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.2 & 0.3 & 0.5 \\ 0.4 & 0.1 & 0.5 \end{pmatrix}$$

用这个模型，我们可以计算出两日后下雨的概率。当然，有好几种途径可以得到下雨的结果，所以我们需要把所有可能的概率都考虑进来。它可能是SSR，可能是SCR，还可能是SRR：

$$\text{SSR的概率} = \text{SS的概率} \times \text{SR的概率} = 0.6 \times 0.1 = 0.06$$

$$\text{SCR的概率} = \text{SC的概率} \times \text{CR的概率} = 0.3 \times 0.5 = 0.15$$

$$\text{SRR的概率} = \text{SR的概率} \times \text{RR的概率} = 0.1 \times 0.5 = 0.05$$

这意味着两日后下雨的概率是0.26或26%（我们用  $S \times S = 0.06 + 0.15 + 0.05$  来表示）。

计算两日后下雨的可能性，还有一种简便的方法，就是将我们前面描述概率的矩阵进行平方。



$$\begin{pmatrix} SxS & SxC & SxR \\ CxS & CxC & CxR \\ RxS & RxC & RxR \end{pmatrix} = \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.2 & 0.3 & 0.5 \\ 0.4 & 0.1 & 0.5 \end{pmatrix}^2$$

尽管每天的天气都依赖于前一天的天气，但从长远来看，无论我们的模型是从晴天、雨天还是阴天出发，下雨的概率都会趋向于相同的数值（32.35%）。为了证实这一点，我们可以提高矩阵乘方的指数来进行验证，结果我们会发现每一行都会趋向于相同的概率。因此，即使明天的天气依赖于今天的天气，长期的天气预报也与今天的天气无关。

下面这个矩阵每一行表示的是十天后的晴天、阴天或雨天的概率。通过数据我们可知，今天的天气是什么并不重要（即我们选择哪一行来观察），第十天的概率总是相同的。马尔可夫设计的这个试验，彻底证明了涅克拉索夫的观点是错误的——从长期犯罪统计数据来看，罪犯决意犯罪是由个人自由意志决定的，这个论点是站不住脚的。

$$\begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.2 & 0.3 & 0.5 \\ 0.4 & 0.1 & 0.5 \end{pmatrix}^{10} = \begin{pmatrix} 0.4412 & 0.2353 & 0.3235 \\ 0.4412 & 0.2353 & 0.3235 \\ 0.4412 & 0.2353 & 0.3235 \end{pmatrix}$$

马尔可夫决定使用一首在俄国家喻户晓的著名诗歌——普希金（Pushkin）的《尤金·奥涅金》（Eugene Onegin）来阐释他的模型。他所做的分析并不是要给这首诗提供新的文学见解，而只是用它作为一个数据集来分析元音和辅音的出现概率。他提取了这首诗的前20 000个字母（大约占整首诗的1/8），计算了元音和辅音出现的次数。虽然计算机可以在一瞬间完成这项工作，但是马尔可夫坐下来，自己一个一个地计算元音和辅音出现的次数。他最终得出的结论是，元音占总数的43%，辅音占总数的57%。如果随机提取一个字母，那么很大可能它是一个辅音。他感兴趣的是，知道前一个字母是辅音还是元音是否会改变你对当前字母的猜测。换句话说，下一个字母是辅音的概率是否取决于前一个字母是不是辅音？

经过对文本的分析，马尔可夫发现，一个辅音后面会跟着另一个辅音的概率是34%，而辅音后面会跟着一个元音的概率是66%。尽管前几个字母在很大程度上取决于起始字母的选择，但是马尔可夫证明，从长远来看，字母的分布是一种模式。因此，即使是相互依赖的事件，如果它们受到固定概率的影响，也是一致的。这其实并不令人意外，因为大多数单词的拼写往往都是辅音和元音交替出现的。经过统计计算，他发现元音后接元音的概率只有13%。因此，可以说《尤金·奥涅金》为马尔可夫提供了一个完整的模型，帮助他阐释了自己的想法。

马尔可夫链的一个重要性质就是其是无记忆的：在当前状态下，你可能需要一切可用的事件来预测下一个事件。这种无记忆性也叫马尔可夫属性。有时可以通过考虑前两种状态如何影响下一种状态来改进模型。（以《尤金·奥涅金》试验为例，了解前两个语音元素，可能有助于增加推测下一个语音元素的正确概率。）但在某种程度上，这种依赖性是不存在的。

## “续作者”：第一个人工智能即兴爵士作曲演奏者

帕切特决定用帕克的作品来替代普希金的诗。他的想法是，对爵士音乐家的即兴演奏片段进行分析，给定一个音符，分析下一个音符出现的概率。现在，让我们想象一个由上行和下行音阶组成的即兴重复乐段。如果演奏一个特定的音符，那么下一个音符上行或是下行的概率都是50%。基于这样的事实，该算法将在音阶上进行随机抓取。给予算法的即兴演奏次数越多，它分析的数据就越多，某一种特定的演奏风格就会出现越多。帕切特发现，仅仅回溯一个音符是不够的，可能需要几个音符才能知道下一个音符是什么。但是，我们不希望算法生成与训练数据雷同的东西，所以回溯过多是没有好处的。

帕切特算法的优点是你可以向其提供实时数据。比如，你可以在钢琴上即兴弹奏，该算法会统计分析你在做什么，当你停下来的那一刻，它会继续以同样的风格进行弹奏。这种问句和答句的形式在爵士乐中很常见，因此该算法可以和音乐家进行现场的旋律问答。正因为该算法可以生成与训练数据相同风格的输出内容，所以被称为“续作者”。

“续作者”会根据它刚刚演奏的内容以及训练数据，来计算下一个特定音符出现的概率，然后以掷骰子的随机模式做出选择。在另一个版本的算法中，帕切特没有选择问句与答句的“问答模式”，而采用了一种“合作模式”：算法会根据演奏的一段旋律，利用它的概率演算来推测出正确的伴奏和弦，就像真人在伴奏一样。

使用该算法的爵士音乐家对其有什么样的看法呢？伯纳德·鲁巴特（Bernard Lubat）是一位当代爵士音乐家，他对“续作者”进行了测试，该算法给他留下了深刻的印象：“这个系统向我展示了一些我本可以实现的想法，但它们需要我花费很多年时间才能真正实现。虽然它走在我前面，但毫无疑问它所发挥的一切的源头还是‘我’。”“续作者”已经掌握了鲁巴特的音乐世界，但它不是简单地重复他以前做过的事情，而是在探索新的领域。可以说，它是一个展示探索性创造力的算法。除此之外，它还通过向鲁巴特展示他以前从未接触过的方面，来促使其写出的作品更具创造性。

对我来说，这就是通过洛夫莱斯测试的时刻，是AlphaGo与李世石制胜一招的音乐版。这个算法生成的结果让编写算法的程序员和对算法

进行训练的音乐家都感到惊讶，但它所带来的意义不仅仅是创新和令人惊讶而已。对于鲁巴特来说，这个算法生成的输出内容太有价值了：在算法帮助下，他的作品更具创造性，他也变得更富创造力了。

人们都倾向于安于现状、故步自封。“续作者”已经开启了新的音乐世界的大门，它强有力地告诉大家：“嘿，你知道吗？你也可以做得得到！”鲁巴特这样说道：“这套系统生成和‘演奏’的音乐，尤其是那些余音绕梁的旋律，以及令人难以置信的节奏，都是人类能力极限的巅峰之作。这也许会改变人们对精湛的艺术的概念。”

鲁巴特觉得在某种程度上他受到了身体的限制，而“续作者”却没有，这使得“续作者”有可能比他更有创造力。通常，缺乏实施的方式阻碍了计算机创造力的表现，但在这个试验中，我们确实实地感受到了计算机的创造力。与人类相比，计算机能够更快地完成任务，处理更多的数据，但这一事实可能会导致人类创造力和人工智能创造力之间有趣的紧张关系。电影《她》（Her），讲述了一个男人爱上人工智能系统OS1的化身萨曼莎的故事，这部电影就在暗示这样一种关系。在影片中，经过一段时间的相处，人工智能OS1学会了与人类交流沟通。在学习的过程中，它与“哲学家”<sup>[1]</sup>建立了更有价值的关系，加之它认为与人类进行互动终究是太慢了（相较于它的CPU运算速度），所以最终选择了离开。也许在未来，“续作者”将生成只有另一台机器才能欣赏到的“声音”，因为它的复杂性和速度超出了人类的极限。

在帕切特设计的一场试验中，“续作者”引起了听众们有趣的情感反应。在现场表演中，尽管鲁巴特在一旁干扰，但听众的反应是惊讶，持续的惊讶，且他们努力地跟随“续作者”的演奏。之后，帕切特决定让“续作者”进行爵士音乐版的图灵测试。他请了两位爵士音乐评论家来进行评判，他们听到的是由爵士钢琴家阿尔伯特·范·维恩达尔（Albert van Veenendaal）和“续作者”以“问答模式”演奏的即兴爵士乐。两位评论家都觉得难以区分，不知道哪一部分是机器演奏的，哪一部分是人演奏的。同时，他们都倾向于认为“续作者”演奏的部分很可能是人类爵士音乐家演奏的。这是因为“续作者”正在以有趣的方式，在体裁、样式、风格方面不断进行着突破。

“续作者”的突破，完成了很多了不起的事情。可是，虽然它创作的音乐片段在局部是有意义的，甚至是相当令人惊喜的，但基于马尔可夫链系统存在的一些内在局限性，令它的作品最终还是不那么令人满意。因为这些作品没有总体的全局结构，也就是并没有实现真正意义上

的“作曲”。帕切特意识到，如果要使音乐具有更强的叙事性，那么就必须驱动、限制旋律进行演进。在问答中，你通常希望从问题开始处得到最终的答案，类似地，在音乐中你会期盼旋律从不稳定解决到稳定，最终实现某种张力的解决。帕切特必须找到一种新的方法，将马尔可夫链系统中的自由与约束条件结合起来，从而形成一种更具结构化的组合。

[1] 电影中以英国哲学家阿兰·瓦茨为原型建造的高级人工智能系统。  
——译者注

## “心流机”

许多艺术家和表演者曾描述，当他们全身心地投入到艺术中时，就失去了时空感，有些人称之为“进入状态”。美籍匈牙利心理学家米哈里·契克森米哈赖（Mihaly Csikszentmihalyi）于1990年首次提出并确立了“心流”<sup>[1]</sup>（flow）这一概念。帕切特决定尝试创造一种算法，来帮助有创造力的艺术家进入“心流”状态。

要想达到“心流”状态，必须在任务的难度和操作者的技术能力之间建立起平衡。如果这两者没有达到平衡，你就会出现图12-1所示的心理状态：如果操作者技术能力不足，任务难度又相对较高，那么他就会陷入焦虑的状态；如果任务难度过低，所需技术能力要求也很低，就会给人带来无聊的感觉。



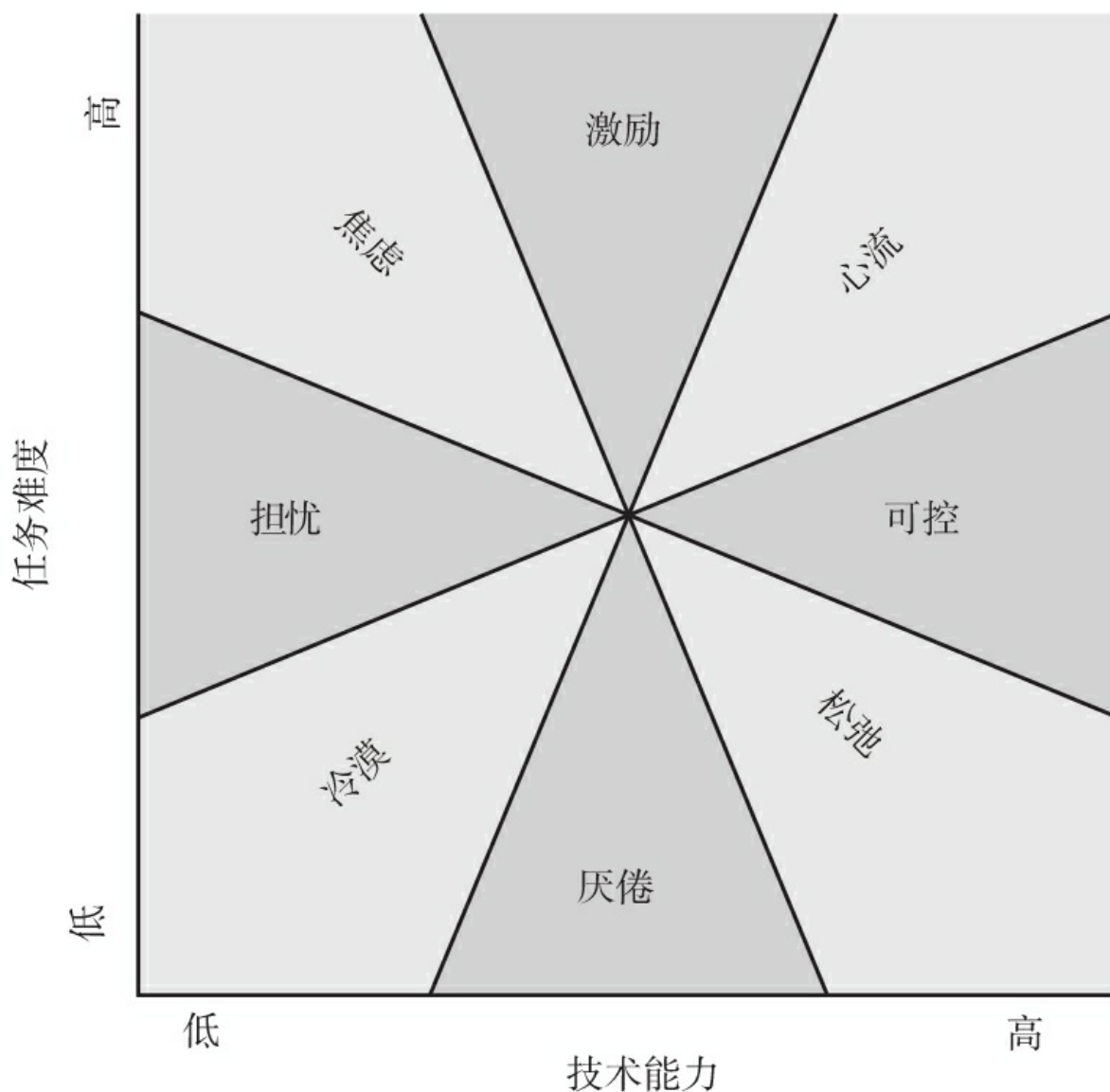


图 12-1

帕切特“心流机”的核心算法是，先使用马尔可夫链来学习艺术家的风格，然后再添加一定的约束条件。许多具有创造力的艺术家也是这样工作的。毕加索花了数年来汲取埃尔·格列柯<sup>[2]</sup>（El Greco）、雷诺阿<sup>[3]</sup>（Renoir）、委拉斯贵支<sup>[4]</sup>（Velázquez）和马奈<sup>[5]</sup>（Manet）的作品风格，通过模仿、组合和调整他们的风格，并添加不同的约束条件，来创造一种属于他自己的独特风格，但这种风格植根于过去的这些大师。

帕切特尝试让他的算法在一种风格下运行，同时从另一种风格中提

取约束条件。这是一个使用博登组合型创造力概念进行算法试验的绝佳例子。在一次试验中，帕切特借鉴了查理·帕克<sup>[6]</sup>（Charlie Parker）的蓝调风格，并将其与皮埃尔·布列兹（Pierre Boulez）的序列音乐的约束条件结合起来（布列兹十分崇拜勋伯格，所以他的作品使用了十二音体系）。这个约束条件就迫使蓝调音乐演奏时要使用十二音循环，这是一个很不一般的约束条件，因为蓝调音乐通常使用三音列作为其发展基础。试验的结果是生成了一个奇奇怪怪的东西，虽明显有查理·帕克比波普爵士乐<sup>[7]</sup>的风格，但总是游离在风格的边缘。我必须承认我很享受这种融合。再用其他人的风格和约束条件来做试验，效果就不是那么好了，比如使用约翰·柯川<sup>[8]</sup>（John Coltrane）的音乐风格并以理查德·瓦格纳（Richard Wagner）的和弦使用为约束条件来演奏《巨人舞步》（Giant Steps），这个效果是难以令人满意的。

“心流机”并不局限于音乐。你可以让它学习一个诗人的风格，并使用另一个人的风格为约束条件。例如，帕切特的团队用马尔可夫模型来学习鲍勃·迪伦（Bob Dylan）歌词的风格，然后将其应用到甲壳虫乐队的《昨日》（Yesterday）歌词中。甲壳虫乐队的歌词在节奏和韵律上有一定的限制，“心流机”的任务是在这个框架中填充被识别认可为鲍勃·迪伦所作的词句。下面就是结果，你可以使用《昨日》的曲调来唱唱看。

### 【原文】

Innocence of a story I could leave today

When I go down on my hands and pray

She knocked upon it anyway

Paradise in the dark side of love it is a sin

And I am getting weary looking in

Their promises of paradise

Now I want to know you would be spared this day

Wind is blowing in the light in your alleyway

Innocence in the wind it whispers to the day

Out the door but I could leave today

She knocked upon it anyway

在此之后，帕切特用“心流机”完成了第一首人工智能创作的流行歌曲。经过多年的研究，他终于实现了儿时的梦想。这首由“心流机”作曲的新歌叫作《老爸的车》（Daddy's Car），采用了帕切特最喜欢的甲壳虫乐队的音乐风格。许多音乐分析家认为，甲壳虫乐队的音乐有一个暗含的模式，帕切特希望破解他们的密码。这首歌的歌词并不是由算法生成的，而是由贝诺·卡里（Benoît Carré）创作的，他还负责将算法生成的各个部分进行混轨合成。

继《老爸的车》之后，帕切特在2018年初发行了专辑《Hello World》。专辑名来自所有人学习编程的第一个练习：创建一个输出文本“Hello World”的程序。这张专辑是卡里和其他一些使用“心流机”的音乐家合作完成的，“心流机”帮助这些音乐家不断地拓展自己的创造力。定义这是人工智能制作的第一张专辑并不十分准确，因为卡里和他的合作者在确定最终产品的边界上发挥了重要的、无可替代的作用。

那么，人们对这张专辑的评价如何？作曲家法蒂玛·卡迪里（Fatima Al Qadiri）很轻蔑地打趣道：“这首歌就好像是一首歌叠置了50遍的结果。”

但并不是所有人都持消极的态度。帕切特被声田（Spotify）从索尼实验室（Sony Labs）挖过来。有消息称，声田正在创建一个全是“冒牌艺人”歌曲的播放列表，这一举动很有意思。乐评家们在声田上发现了一些点播量超高的艺人，他们的作品被列入冥想或慢跑的流行播放列表。一个名为“深望”（Deep Watch）的乐队在5个月内得到了450万次的点播量。

当人们试图找出这些艺人而在上网搜索时，却总是搜索不到任何信息。没有演唱会的通告，也没有乐队、乐手的简介，总之没有任何信息。有传言称，这些音乐是由“冒牌艺人”创作的，这样声田就不需要支付版税了。每每这时，声田就会这样打官腔：“我们过去没有，未来也不会创造‘冒牌艺人’，并把他们的作品放在我们公司的播放列表里。你们讲的全是子虚乌有！完毕！”但似乎他们是专门委托名不见经传的小

艺人以假名创作歌曲，这样他们需要支付的版税就要比与唱片公司签订的标准协议优惠得多。

事实上，这些艺人有可能写出无穷无尽的“口水歌”，这就是公式化创作的结果，令人惊讶吧！与古典经典作品不同的微妙之处在于，许多流行歌曲只是在重复已验证可行的格式，而没有考虑去挑起人们的期许和愿望。这样的流行歌曲一般是4/4拍，4小节或是8小节一个乐句，旋律不停地一遍一遍重复。这样做会让这首歌很容易上口，但是会陷入死循环，不会有新的旋律。当然，也会有令人耳目一新的“昙花一现”，但往往最后的结果是创造了一个新的格式，歌曲又陷入了一句旋律一遍一遍重复的死循环。

声田雇用帕切特会不会提升竞争的水平？会不会导致这些艺人失业？算法已经开始控制我们所听的内容了。还要多久，算法才能为我们定制歌曲？如果实现了，声田就不用再支付任何版税了，只需要发薪水给帕切特就好。

如果想要一段个人定制的人工智能生成的音乐，你可以访问Jukedeck网站。Jukedeck是由两名剑桥大学的毕业生创建的，他们是发小，8岁时就在教会唱诗班认识了。Jukedeck是众多利用人工智能为机构和公司创作歌曲的公司之一。他们的顾客，从自然历史博物馆这样的机构到可口可乐这样的公司都有。这些公司需要原创但便宜的背景音乐来制作视频和广告，它们不想支付高昂的版税。Jukedeck利用人工智能几秒钟就可以完美地为视频配乐。

这个网站提供了不同类型的音乐，任你选择。从阳春白雪到下里巴人，从合奏到鼓和低音。然后，你需要告诉它你想要的音乐是积极的、忧郁的，还是其他八种情绪中的哪一种。在你选择并确认后，这个算法就会生成时长90秒的音乐，它甚至会为这个音乐起个不错的名字。

我的选择是为科幻电影配乐，算法生成了一首名为Impossible Doubts的音乐。这段音乐不会吸引我常常去听，但这并不重要。“足够用了”这句话在人工智能音乐时代被经常使用。Jukedeck的目标是创作视频制作或游戏开发中所需的背景音乐，而不是与阿黛尔打擂台——有了一种能够对情绪做出反应的算法，就有了一个完美的工具，它可以追踪玩家在游戏轨迹。如果想听《Impossible Doubts》，我可以花0.99美元获得免版税的使用许可，或者花199美元购买版权，直接买下这首歌。

也许这样的价格是人工智能音乐背后一个重要的驱动力。金钱推动着人工智能的艺术革命而不是对艺术方面的考量。

[1] 在《心流：最佳体验的心理学》一书中，米哈里·契克森米哈赖概述了此理论：人们在心流状态下最为快乐，这是一种对正在进行的活动和所在情境的完全投入和集中，是一种人们因为过于沉浸在一项活动中而忽略身边一切事物的状态。心流的概念与处于最佳状态的感觉一致。心流状态是内在动机的最佳形式，在这里人可以完全沉浸在他所做的事情中。这是一种人人都会有的感受，它是一种强大的吸引、投入、满足和熟练，而在此期间暂时性的关注（时间、食物、自我等）一般都会被忽略。——译者注

[2] 西班牙文艺复兴时期著名的幻想风格主义画家。他的作品构图奇特，布局多呈幻想结构，用色大胆、新奇，呈现出梦幻般的奇特效果。——译者注

[3] 法国画家，最初与印象画派运动联系密切。他的早期作品是典型的记录真实生活的印象派作品，充满了夺目的光彩。18世纪80年代中期，他从印象派运动中分裂出来，转向人像画及肖像画，在妇女肖像画中发挥了自己更加严谨和正规的绘画技法。——译者注

[4] 巴洛克时期西班牙画家，是巴洛克风格大兴其道时代的一个伟大的写实主义者。他的画尤为重视色彩表现，他使用复杂色彩的技巧超过了前辈大师。如果仅从色彩上讲，委拉斯贵支是当之无愧的巴洛克大师，因为他的色彩理念淳朴自然。——译者注

[5] 法国著名印象派画家，第一位把印象主义的光和色彩带进人物画的画家，开创了印象主义画风（请注意，不是莫奈）。——译者注

[6] 中音萨克斯演奏家，1955年去世，对比波普爵士乐的贡献最大。——译者注

[7] 比波普爵士乐，亦称Bop，一种激进的爵士乐风格，可看作现代爵士乐的前身。出现于20世纪40年代中期，由爵士小号手约翰·伯克斯、古莱斯皮和萨克斯手查理·帕克共同创立。——译者注

[8] 1926—1967，萨克斯表演家和作曲家，爵士乐大师，自由爵士乐精神之父。——译者注

## 量子作曲

艺术创作的一个奇特之处是，艺术家创作的作品必须吸引许多不同的人来观看、阅读或聆听。但每一个受众都有不同的品味、期望和情绪。如果你能创造出颠覆这种观念的艺术，即转而寻求许多的作品针对一个人，那会怎样呢？我们的智能手机收集了大量关于我们的数据，如果所有这些信息都可以用来为我们量身定制一件艺术品呢？

“大举进攻”乐队正在这样做。自2010年发行专辑Heligoland以来，该乐队没有发布任何新专辑，直到2016年年初他们选择了一种创新的方式发布了四首新歌。粉丝们通过下载、安装并运行一款名为Fantom的应用程序来收听歌曲，这个应用程序是为这四首作品专门开发的。一旦你允许应用程序访问你的位置信息、时间信息、摄像头图像、心率和Twitter，算法就会决定如何为你现场混编音乐。

“大举进攻”乐队的算法本质上就是高级复杂版的莫扎特骰子游戏。原来的音轨被分解成更小的迷你音轨，作为创建新的个性化音轨的原材料。在新歌发展的每一个阶段，算法都会做出选择：决定接下来要添加哪部分迷你音轨，以及以何种方式混合。这些决策由算法从用户处收集的数据来决定。用户的心率、动作、摄像头捕捉到的画面……都将影响你所听到的歌曲的旋律和音质。

该算法的关键之处在于创建了一棵可能树，源树提供的可能性足够丰富和多样，又具有足够的连贯性，所以无论算法选择哪条路径，结果都会显得天衣无缝、那么自然。相信我，你想要的绝不是完全的随机性。莫扎特精心策划了每一个小节，提供了11个选项，每一个都可以作为华尔兹舞曲的下一个小节，而华尔兹的整体结构确立了游戏规则。“大举进攻”乐队的算法亦是如此。这样才能保证在音乐发展的过程中不会出现和声的冲突和崩塌。

创建这款应用程序的程序员罗布·托马斯（Rob Thomas）相当贴切地将其称为“量子作曲”。在量子世界里，一个电子因量子叠加的缘故，可以同时出现在许多不同的地方。是观察的行为导致了波函数坍缩，坍缩成它的许多种可能状态之一。托马斯的想法是创作一首可以存在于许多种可能状态之中的歌曲。那么，当我决定听这首歌的时候，算法会根据我的数据和做出的选择，将“大举进攻”的“波函数”坍缩成一首歌曲。



托马斯感兴趣的是我们的情感状态和我们所听的音乐之间的关系和相互之间的影响。他说：“音乐是一种情绪处理器。我想知道如何利用音乐技巧来诱导听众的情绪状态。”目前，他正在探索开发使用人工智能音乐帮助诱导人进入冥想状态的应用程序。他的设想是这样的，音乐对人当前的精神和身体状态的数据做出判断和反应，让算法学习如何操纵人的精神和身体来使之放松。当然，托马斯也承认，如果想驾驭最有效的情绪处理器，那么你真的需要创造一个人。

应用程序Fantom依赖于音乐家对歌曲各部分的管控能力。“大举进攻”乐队认识到了机器学习的强大功能，所以他们能够以一种更有机的方式创建可能选择的树结构。该乐队希望，在下一个版本中让机器学习创造自己版本的音轨。罗布·托马斯与伦敦大学金史密斯学院（Goldsmiths University of London）的米克·格里森（Mick Grierson）合作，初步实现了这一目标。

格里森曾与冰岛先锋乐队西格尔·罗斯（Sigur Rós）密切合作。他把他们的一首歌“风暴”（Óveður）扩展成一个24小时的版本，在其中不会有任何重复，但保留了5分钟的原曲目。这个时长24小时的版本是为了配合一次环绕冰岛海岸的旅行而制作的，这次旅行在YouTube和冰岛国家电视台上进行了直播和转播。作为“慢电视”<sup>[1]</sup>（Slow TV）新热潮的一部分，2016年6月20日这一旅程开始了。艺术家沿冰岛海岸自西向东前行了1332公里，经过了欧洲最大的冰原、冰川泻湖、东海湾和莫兹勒达勒荒凉的黑沙滩。

对于一个人类作曲家来说，创作一个24小时不重复的原声音乐是相当困难和耗时耗力的。格里森开发的软件使用概率工具生成音轨，来响应音乐所附的图像。后来，他还创作了这首歌的一个更长的版本，这个版本将永远播放，永不重复。即使“大举进攻”乐队或是“西格尔·罗斯”乐队解散了，借助这个算法，只要我们想听，还是会有源源不断生成的新版本的歌曲。

布莱恩·埃诺（Brian Eno）创造了“生成音乐”（generative music）一词，来描述由一个系统或算法创造的不断变化的音乐。埃诺这样说：“它是为自己思考的音乐。”这是一种音乐苗圃，作曲家播种下种子，算法与外界互动（如一个人玩电脑游戏，或者其一天的经历），然后这些种子就会生长出声音。就某种程度而言，现场表演意在捕捉这样一种理念：从乐谱到体验。但在每一次演绎中都会产生一些独特的东西。埃诺对进一步推进这个想法很感兴趣。他的应用程序，比如Bloom

和Scape，或是他与彼得·奇尔弗斯（Peter Chilvers）合作开发的Reflection，会产生无穷无尽的“类埃诺”的音乐，这些音乐是用户通过智能手机与程序交互而生成的。他这样描述生成的过程：“河，还是那条河，时时刻刻变化着。”

虽然埃诺在他的创作中融入了技术，但他和洛夫莱斯一样不相信他所使用的算法会生成任何超出创造者所输入的东西：“在众多的含义之中，我们已经做出了很多审美的选择。当有人利用程序创作了一段音乐，他们就是在与我们合作创作它。”

机器学习正在挑战人类作曲家所依仗的洛夫莱斯对人工智能的预言。2016年，一个名为AIVA的算法成为被法国音乐人、创作人和版权代理商协会（简称SACEM）授予作曲家称号的第一台机器。该算法由皮埃尔·巴罗（Pierre Barreau）和文森特·巴罗（Vincent Barreau）两兄弟创建，该算法学习了巴赫、贝多芬、莫扎特以及其他海量的音乐家的作品乐谱，产生了一位正在创作自己独特音乐的人工智能作曲家。尽管目前正在为电脑游戏创作音乐，但它的目标是崇高的——“要在音乐史上留下浓墨重彩的一笔”。听了AIVA的第一张专辑《创世纪》后，我觉得巴赫和贝多芬还不用太担心。但正如专辑名所示，这只是音乐人工智能革命的开始。

**[1]** 正在北欧国家流行的慢节奏直播电视节目，以慢节奏、超乏味的特点吸引观众注意。不像传统的电视节目，其没有快速的剪辑镜头，没有精美的后期制作，是一种全新的看似“古怪”的电视节目类型。——译者注

## 人为何创作音乐

音乐一直具有算法性质，这意味着在所有的艺术形式中，它受到人工智能进步的威胁最大。音乐也是所有艺术形式中最抽象的一种，它利用结构和模式，而正是这种抽象的性质使它与数学紧密相连。但这也意味着，在这个世界里，算法将像人类一样得心应手。

但音乐不仅仅是结构和模式而已，必须通过表演来赋予其生气。在文明初始，音乐用于配合特定的宗教仪式。在我们祖先画满壁画的洞穴里，考古学家发现了乐器存在的证据：秃鹫骨头制成的骨笛、动物角制成的号角以及原始的弦乐器。

一些人推测这些原始的工具可能是用来交流的，但另一些人认为它们是我们祖先早期发展的宗教仪式的重要组成部分。似乎对宗教仪式的需求是人类代码的一部分。宗教仪式由一系列的活动组成，包括手势、言语和物品摆放，这些活动在一个神圣的地方按照固定的顺序或模式进行。通常，从外部来看，仪式似乎是非理性或不合逻辑的，但对内部来说，它提供了一种联系团队的重要方式。音乐在许多这样的宗教仪式中扮演着重要的角色。在唱诗班唱歌或在乐队中演奏是一种将不同的意识体验融合在一起的非凡方式，就像在体育比赛中，我们在看台上一齐演唱的歌曲将我们团结在一起，共同对抗客场的球迷。

当智人迁移到欧洲并遇到尼安德特人时，音乐的这种结合群体的能力可能给了智人以优势。正如作曲家马尔科姆·阿诺德所写的：“音乐是人与人之间交流的社会行为，是友谊的象征，是强大无敌的。在德国发现的4万年前旧石器时代的笛子表明，我们的祖先可能能够进行远距离的交流。”人们很快意识到，在创造改变思维的仪式时，音乐是一个强有力的因素。不断重复可以帮助改变我们的意识状态，正如许多萨满教仪式所证实的一样。我们的大脑具有与不同的心理状态相对应的固有频率。恍惚音乐充分利用了这个事实，即每分钟120拍的音乐最适合诱导人类产生幻觉体验。从现代试验中我们得知，混淆多种感官输入会导致大脑产生奇怪的灵魂出窍的体验。例如，触觉和视觉的组合可能使人产生肢体识别的障碍或假象。这就是为什么我们经常能在发现早期的乐器时一同发现香料或草药，因为它们的组合既能给仪式带来一种气味（嗅觉），也能给仪式带来一种声音（听觉）。对一个没有体验的算法，怎么能期望它可以理解音乐的力量，然后来影响甚至改变我们的身体和思

想呢？

随着文明的发展，音乐继续是仪式世界中的一部分。从帕勒斯特里纳到巴赫再到莫扎特，音乐的巨大进步往往是在宗教背景下取得的。有人猜测，随着我们内在意识的觉醒，“上帝”的概念将出现在人类身上：随着意识的发展，意识到自己头脑中有一个声音，此时你会感到震惊。这一定是很可怕的。我想强调的是，宗教仪式和音乐可以安抚大脑中的声音，而自然的力量似乎是众神所为。

这一切听起来都与计算机逻辑和其无情感的世界相去甚远。但算法已经学会了生成让我们感动的声音：Algoraves现在使用的算法可以对跳动的人群做出反应，帮助DJ策划舞曲和打碟；“深度巴赫”正在为教堂唱诗班创作更多的宗教合唱来歌颂上帝。尽管这些算法似乎已经破解了音乐的代码，但是机器内部仍然是毫无波澜的。这一切就像现代的数字“吼板”<sup>[1]</sup>，它们仍然只是我们的工具而已。

**[1]** 吼板是一种自旧石器时代流传至今的古老发声器，在世界许多文化中都能找到其踪迹：一臂长的坚硬木板被固定在一根长绳的末端，只要持有者将它甩过头顶并不断旋转，它就会发出一种360度环绕音响效果的轰鸣声，这会给持有者带来迷幻感，而声音可传至很远的地方。——译者注



## 第13章 深度数学

保尔·瓦雷里（Paul Valéry）

组合与选择，是发明新事物的两个不可或缺的条件。

在英国皇家学会的一次讨论机器学习如何影响未来的会议上，我凑巧坐在戴密斯·哈萨比斯旁边。我们俩最近都成了英国皇家学会的研究员，这是科学家至高无上的荣誉之一。哈萨比斯的AlphaGo引发了我对生存危机的思考：数学家在未来世界是否还有存在的必要？所以我脑子里就萌生出一个想法：“倘若哈萨比斯能让算法打败世界顶级的九段高手，难道他就不能开发一个证明数学定理的算法吗？”这样的算法也足以让他成为英国皇家学会的研究员。

我向哈萨比斯提出这个疑问后，他的回答令我大吃一惊。“我们已经着手在做这件事情了。”他低声对我说。好像什么事情都逃不脱他们敏锐的洞察力。会议结束后他向我做了详细的解释：他们已经组建了一支团队，目的是利用以往的数学定理证明过程训练算法，以发现新的定理。哈萨比斯邀请我到位于潘克拉斯广场6号的DeepMind研发中心参观，以便了解他们最新的进展情况。

带着些许惶恐，我开始着手研究人工智能时代数学与机器学习之间的博弈。虽然DeepMind在2014年就被谷歌以4亿英镑的价格收购了，但在哈萨比斯的坚持之下，该团队依然留在了伦敦，位于国王十字车站旁边的谷歌伦敦总部内。一进车站大厅，我就看到一大群人在9¾站台（电影《哈利·波特》中通往魔法世界的入口）前排队等候拍照。那一幕触动了我，我觉得体验真正魔法的地方不应该在这里，而是隔壁。

谷歌设在伦敦的总部整体有一种现代牛津大学的感觉，提供了有助于员工们集中注意力、进行深度思考的最佳设施及环境：24小时免费食物供应，配有专门的咖啡师随时为激活员工大脑活力而服务；90米长的跑道，提供免费按摩服务；甚至还可以上厨师丹·巴滕（Dan Batten，曾与英国厨神杰米·奥利弗共事）的烹饪课。当大脑处于超负荷时，员工还能去遍布于大楼各处的“睡眠仓”里美美地睡上一觉。

这还只是谷歌设在伦敦的临时办公场所，一旁高端大气的新总部正

在修建之中。这座气势非凡的建筑由丹麦建筑师比雅克·英格尔斯（Bjarke Ingels）和2012年伦敦奥运会主火炬的设计者——英国设计师托马斯·赫斯维克（Thomas Heatherwick）联合设计，被称为“地面雕刻机”（landscraper）。其主体共11层，长330米，就像一个巨大的脚印。如果立起来，它比伦敦最高的碎片大厦（The Shard）还要高。

“不会休息，就不会工作”，谷歌深谙此道。谷歌富有创意又极具人性化的办公室设计享誉全球，是无数程序员心中的天堂：维多利亚办公区有一间乐器房，供员工们在休息时间尽情演奏，享受音乐带来的放松及愉悦；加州山景城办公区拥有独立的保龄球馆。但这些跟位于国王十字车站的新总部相比，可谓是小巫见大巫——它拥有奥运会规格的游泳池和令人惊叹的屋顶花园。开阔的屋顶花园是这栋建筑的一大亮点，它是一个300米长的多层空中花园，每一层有不同的主题，“高原”“花园”“田野”，种植有草莓、醋栗和鼠尾草。花园被划分出不同的区域，比如植物繁盛的休息区、咖啡区，甚至有长约200米的跑道，等等。员工在工作间歇可以尽情娱乐放松，如果愿意的话还可以在那里编写代码。谷歌办公场所的“豪华”正是机器学习蓬勃发展的明显象征。

DeepMind占据了总部大楼两层的空间，其中一层用于商业应用，另一层用于研发。位于6楼的研发部正在开展的一系列有趣的项目迅速吸引了我的注意：机器学习正用于帮助人们探索量子物理这个难以捉摸的随机世界，同时，其也通过各种各样的项目逐渐渗透到生物学和化学领域。但我的兴趣点在于数学，我很想了解一下他们在这个领域做了哪些工作。

哈萨比斯建议我和奥利奥尔·温雅尔斯（Oriol Vinyals）谈谈，了解一下他们在数学证明方面的最新进展。温雅尔斯是西班牙人，本科时在西班牙学习数学，但他很快发现自己更热爱人工智能。他曾前往加州攻读博士学位，也正是在那里，他被Google Brain选中，后来又转入DeepMind。

当电梯门打开，温雅尔斯向我打招呼时，我既紧张又兴奋，但很快就放松了下来。Google的工作氛围提倡随性而平等，你甚至可以穿T恤和牛仔裤来上班。它的随性自然不仅指着装，更主要是人与人之间可以坦诚相见、畅所欲言。

所有的会议室都是以阿达·洛夫莱斯等科学先驱的名字命名的，我们选中了其中一间。温雅尔斯说：“我们所从事的这项数学研究不仅有



DeepMind参与，还有遍布世界各地的谷歌研究团队的成员。”值得这么多谷歌人探索的是什么样的数学呢？他们的兴趣点会是我所研究的对称世界中的某个定理吗？还是与图论和组合学相关的证明？抑或是确定费马大定理的变体是否有解？温雅尔斯很快透露，他们将从一个截然不同的角度（我认为跟我所了解的数学完全不相关的角度）来研究一个我曾预见的问题。

## Mizar的数学

DeepMind和谷歌研究团队的关注点在于一个20世纪70年代在波兰启动的名为Mizar的项目。该项目旨在构建用一种容易被计算机理解和检验的形式语言描述的数学证明数据库系统。波兰数学家安杰伊·特里布里克（Andrzej Trybulec）率先启动了该项目的研究。Mizar的名字源于大熊座中的一颗恒星——开阳星，这个名字是由特里布里克的妻子取的——当被丈夫询问能否给他的新系统取个好听的名字时，她恰好在翻看一本天文图集。

该系统允许任何人提交用这种形式语言编写的数学证明。因此，到2013年特里布里克去世时，Mizar已成为世界上最大的计算机数学证明数据库：其中一部分是将人类证明过程转化为计算机语言，另一部分则由计算机直接生成。该系统目前由波兰比亚威斯托克大学、加拿大阿尔伯塔大学、日本信州大学的研究小组负责开发和维护。近年来，人们对该系统的关注程度有所下降，数据库的发展不是很快。DeepMind和谷歌研究团队将其目标锁定在Mizar的数据库上，这一点超出了绝大多数人的预料。

经过数十年的积累，人们已经用形式语言这种计算机更容易理解的语言在Mizar的数据库中创建了5万多个定理。Mizar项目的参与人从中挑选出许多人类数学家们钟爱的定理，比如代数基本定理：复数域上的 $n$ 次多项式有且仅有 $n$ 个根。

这个定理非常具有代表意义，其证明可谓一波三折。从17世纪初到现在，人类对它的证明中曾出现过多次失误，其中不乏最伟大的数学家，诸如欧拉、高斯、拉普拉斯等人。直到1806年，让·罗伯特·阿甘（Jean Robert Argan）才提出第一个被公认为完整无误的证明。以往证明中的错误都藏得很深，没有足够的时间检验，是极不容易被发现的。倘若计算机能够发现人类证明中隐藏的错误，那么它在证明定理方面的正确性和有效性就会被刮目相看了。

计算机在Mizar数据库中生成证明的流程是这样的：首先，整理出数学、几何学的基本公理列表；其次，制定推理规则；最后，用一系列相互关联的推理规则构建出某一定理的证明过程。这与下围棋有异曲同工之处：一开始棋盘上空空如也，推理规则就相当于黑白棋手交替行棋

所遵循的游戏规则，定理就相当于最终形成的某种特定的棋局。

DeepMind团队意识到，证明定理和下围棋在本质上是相互关联的：两者都是在可能的输出结果树中寻找特定的节点。而每个节点又具有不同分支，且到达某一特定终点（叶子节点）的分支长度有可能非常长。问题的关键就在于如何选择分支以获得最期待的输出结果：赢得一场比赛或证明一个定理。

利用计算机来生成数学定理已是司空见惯、不足为奇的事情了，甚至略微夸张一点说，计算机只要启动，就可以证明定理。但“条条大路通罗马”，一个定理的不同证明中往往会出现重叠。可这并不是研究人员的兴趣点，他们真正要解决问题是，在被给定一个命题（特定的终点）时，计算机是否能够找到通往该终点的路径，即命题的证明。如果不能，那能说明这个命题是假命题吗？

通过对Mizar相关数据的研究，DeepMind和谷歌研究团队发现其约有56%的定理证明没有人类参与的痕迹。他们将目标锁定在创建一个新的定理证明算法以提高这一比例：用计算机成功生成的证明来训练该机器学习算法，通过对Mizar数据库中已有数据的学习获得探索证明树的好方法。温雅尔斯拿着手中的研究论文无比自豪地对我说，他们的算法已将Mizar数据库中机器证明的比例提高到了59%。这个看似微不足道的“一小步”，代表的却是新技术应用的“一大步”。它不仅仅是多证明一个定理或者多赢一场比赛，而是计算机可以完成的证明量增加了3%。

在某种程度上，我可以理解温雅尔斯的激动之情。这就像学习演奏爵士乐的算法，决定它前景的是一个合乎音乐继续发展的逻辑，而不是接下来到底演奏哪个音。该算法在很大程度上扩展了计算机的应用范围，打开了计算机产生定理的新篇章。

然而，离开DeepMind时我的心情有些失落，因为此行并没有得到预想的收获。我本该为数学能取得如此巨大的进步而高兴，但实际上，机器只是盲目地生成了一些粗制滥造的“数学音乐”，而不是我所期望的“天籁之音”。没有人评判这些新发现的价值，也没有人对其中是否有令人惊讶的启示而感兴趣——它们只是新的而已。

## 数学图灵测试

这就是真正的未来吗？我从Mizar的数据库中精心挑选了一部分定理的证明，想要更深入地了解一下。为了能通过一个真命题推导出另一个命题，证明用计算机形式语言描述。但我却发现自己几乎无法驾驭这种晦涩难懂的形式语言，我体会到了大多数人打开我的论文看到一堆看起来毫无意义的符号时的那种感受。这不是人类表达和交流数学思想的方式。例如，关于“素数有无穷多个”<sup>[1]</sup>这一定理，Mizar的证明过程如下：

---

```
reserve n,p for Nat;
theorem Euclid: ex p st p is prime & p > n proof
set k = n! + 1;
n! > 0 by NEWTON:23;
then n! >= 0 + 1 by NAT1:38; then k >= 1 + 1 by REAL1:55;
then consider p such that
A1: p is prime & p divides k by INT2:48; A2: p <> 0 & p > 1 by A1
thus p is prime by A1;
assume p <= n;
then p divides n! by A2,NATLAT:16;
then p divides 1 by A1,NAT1:57;
hence contradiction by A2,NAT1:54;
end;
theorem p: p is prime is infinite
from Unbounded(Euclid);
```

---

即便是我这样的数学家、专门人士也觉得一头雾水！这不符合人类的叙事方式，甚至可以夸张一点说是语言障碍。

既然通过算法可将西班牙语翻译成英语，那么能不能将这种计算机证明语言翻译成易于与人交流的方式呢？剑桥大学的两位数学家蒂莫西·高尔斯（Timothy Gowers）和莫汉·加内萨林加姆（Mohan Ganesalingam）开展了此项研究。1998年，高尔斯成为菲尔茨奖获得者并登上新闻头条，同年被聘为劳斯·鲍尔（Rouse Ball）讲席教授。

另一位数学家加内萨林加姆的经历也极富传奇色彩。起初，他按部就班地在剑桥大学三一学院学习数学，以第一名的成绩拿到剑桥大学的数学专业学位，并获得资深兰格勒头衔（Senior Wrangler），这是剑桥

数学学子的最高荣誉。后来，加内萨林加姆改行学英语，又以剑桥大学英语学院最佳成绩毕业，获得了盎格鲁-撒克逊英语（Anglo-Saxon English）硕士学位，令所有人大吃一惊。紧接着，他继续攻读计算机科学博士学位，从形式语言学角度对数学语言进行分析。他的数学和语言学的背景很快就派上了用场——高尔斯和加内萨林加姆在三一学院相遇，并惊奇地发现他们对揭开计算机语言难以理解的奥秘有共同的爱好。他们决定一起组建团队，创建一个能够生成人类直接能读得懂的计算机证明。

为测试算法，他们在高尔斯的博客上发布了一个调查问卷。该问卷挑选了本科一年级《度量空间》课程里面的5个定理，每个定理包括3个不同的证明，分别由博士生、本科生和计算机算法完成。为了保证调查的真实性和有效性，博客读者事先未被告知任何关于证明的来源信息。高尔斯仅仅要求他们根据自己的判断为这15个证明的优劣打分，其目的是想了解在没有任何提示的情况下，是否有人会怀疑这些证明不全是由人类完成的。紧接着，他们在第二篇博客文章里向读者透露了每个定理对应的3个证明中有一个是由计算机算法生成的，请读者投票加以辨别。

通过对投票结果的统计分析，大约有50%的读者识别出了由计算机算法生成的证明，但其中只有半数人确信自己的判断是正确的。此外，那些确信不是计算机证明而实际是计算机证明的投票占比也不容忽视。那些来自本科生的证明往往被误认为是计算机的证明。

那么，高尔斯这位菲尔兹奖的获得者是如何看待人工智能“入侵”数学领域对他构成的威胁呢？高尔斯在他的博客中这样写道：

在计算机最终取代人类工作这一历史发展进程中我看不到任何实质性的障碍，这可能会让人感到难过。但实现这一目标的过程却让人憧憬和兴奋。计算机在处理证明中那些烦冗、琐碎环节时的能力越来越强，人机互动越来越少，这留给我们更多的时间和精力去自由地思考更“有趣”的环节。

在Mizar项目中，除了语言问题之外，DeepMind和谷歌研究团队费了九牛二虎之力提高的3%中，有没有出乎意外的“惊喜”呢？我觉得整个项目似乎并没有抓住研究重点。我为什么这样说？

[1] “素数有无穷多个”定理的证明之一：假设素数只有有限的 $n$ 个，其中

最大的素数是 $p$ 。设 $q$ 为所有素数之积加上1，即 $q = (2 \times 3 \times 5 \times \dots \times p) + 1$ ，则 $q$ 不为素数。那么， $q$ 就可以被2、3、... $p$ 中的某一个数整除。而根据公式， $q$ 被2、3、... $p$ 中任意一个数整除后又会余1，与前结论相互矛盾。所以，由此可证明，素数个数是无限的。——译者注



## 巴别数学图书馆

我最喜爱的短篇小说之一，阿根廷作家豪尔赫·路易斯·博尔赫斯（Jorge Luis Borges）创作的《巴别图书馆》可以解释上述的问题。该小说讲述的是一名图书管理员“探索”自己图书馆的故事。小说开头他这样描述自己的工作场所：“宇宙（别人管它叫图书馆）由许多六角形的回廊组成，其数目不能确定，也许是无限的……任何一个六角形回廊的上层和下层看起来都是永无止境的。”除图书馆之外，这里别无他物。它是我们自己的图书馆（我们称之为宇宙）的隐喻。这个像巨大蜂巢一样的图书馆里堆满了大小一致的书籍：每本书有410页，每页有40行，每行由80个书写符号（书写符号共25种，包括空格、句号、逗号以及22个字母）组成。

当图书管理员翻看这些书籍时，他发现除了偶尔地会看到一些有趣的文字之外，几乎所有书籍的内容都是无序、混乱的：有一本书从头至尾全部都在重复MCV三个字母；另一本书则纯粹是完全看不懂的“字母迷宫”，唯有倒数第二页上出现一行字——“啊！时间，你的金字塔！”

图书管理员给自己设定的目标是确定图书馆是否真的是无限大，如果不是，那么它到底是什么形状？随着故事的发展，一个假设被提出：“这个图书馆是‘完全的’……图书馆的书架上收藏着由25个书写符号构成的全部可能的组合（其数目尽管很大，但却是有限的）。换言之，就是能够用所有的语言表达出来的一切。”这个图书馆收藏着有可能被写出来的每一本书籍：托尔斯泰的《战争与和平》随处可见；达尔文的《物种起源》、托尔金的《指环王》，以及这些作品所有语言的译本；甚至本书也被放置在图书馆某个角落的书架上。（到目前为止，我的这本书才写了这么多，我多么希望能够找到它，这样就省得自己苦思冥想剩下的部分了！）

由于所有书的页数、行数、每行的字数这些指标都是固定的，我们可以估算出图书馆的藏书总数。已知构成书籍内容的书写符号共有25种，那么第一页第一行第一个字符就有25种选择，第二个字符也有25种选择，所以前两个字符总共可构成 $25 \times 25 = 25^2$ 种选择。依此类推，每行有80个字符，就有 $25^{80}$ 种可能的组合方式。

我们把问题进一步扩展，计算一下第一页可能有多少种不同的组合

方式。因为每页有40行，所以就有  $(25^{80})^{40} = 25^{80 \times 40}$  种可能的组合方式。每本书有410页，进而可得  $(25^{80 \times 40})^{410} = 25^{80 \times 40 \times 410}$  种可能的组合方式，这就意味着图书馆的藏书总数达到了  $25^{40 \times 80 \times 410}$  本。这个数目非常巨大。给定宇宙可观测范围内的原子总数为  $10^{80}$ ，那么用一个原子代表一本书，即使把所有的原子都用光，也远远达不到巴别图书馆里的藏书总数。但即便如此，它依然是一个有限的数字。根据这个原理，我们可以很容易地编写出程序，让计算机在有限的时间内系统地生成所有书籍。当然，宇宙逐渐衰变成永恒的、冰冷的黑暗要经过多长的时间尚未得知，这里仅仅是从理论上加以讨论。

当人们听说图书馆收集了所有能被写出来的书籍时，首先得到的是一种奇特的幸福感，但随之而来的是巨大的失望，因为人们意识到这个似乎包罗万象的图书馆里实际上什么都没有。托尔斯泰、达尔文、托尔金甚至我的书在出版以后会被牛津大学图书馆收藏，是因为它们被人（许多人）认为是文学世界的瑰宝，它们值得在那里被收藏。这也是巴别图书馆与牛津大学图书馆最大的不同之处。

当我们来到数学书库，会看到那里收藏了《数学年鉴》《l'IHES数学出版物》这些伟大的期刊。那么，要具备什么条件才能成为该书库书架上的一员呢？许多人理所当然地会认为，这个书库一定期望自己能发展成为“巴别数学图书馆”，收录历年来数学家们记录的所有关于数字和几何学的新发现，例如， $\sqrt{2}$  是无理数、有限单群分类列表、球体体积公式、最速落径识别等。

这其实是Mizar想要实现的：首先创建一个数学命题列表，然后用公理去证明这些命题以验证其真假。对命题的证明就是进入Mizar数据库的必要条件。换言之，对于命题的实质是什么，是否有人会觉得它足够有趣，是否可以与其他数学家分享等，Mizar并不关心。它所做的是，只要是对命题的证明，就在没经过筛选的情况下收录到数据库中。换言之，它只是一个包含可以证明的一切的“巴别图书馆”。

在我看来，这违背了数学精神。数学不仅仅是由一组我们所能发现的关于数字的真命题构成的。这可能会让大多数非数学专业人士感到震惊。数学家们像《巴别图书馆》的作者博尔赫斯一样，都是写故事的好手，只不过他们“写作”用的字符是数字和几何图形，而他们证明定理的过程就是在叙述故事和塑造角色。他们判断和选择故事是基于对故事情节产生的情绪反应。

此处引用我的偶像之一、伟大的数学家亨利·庞加莱对数学创造做出的解释：“什么是数学创造？它并不意味着对已知的数学事实重新组合。任何人都可以做到重新组合，但这种组合的数量是无限的，并且大多数毫无价值。创造，意味着不制造无用的组合，而仅制造那些少量且有用的。创造即甄别，即选择。”数学是被创造的还是被发现的？我们之所以认为它是被创造的，归根结底是鉴别和选择。当然，创造方法其他人也可以想得到，但尽管方法很多，却不是人人都能创造出像贝多芬的《大调赋格》（Grosse Fuge）或者艾略特的《荒原》（The Waste Land）那样的伟大作品。数学中也存在着“同样的自由”，这一点可能会令绝大多数人惊讶无比。

正如庞加莱所说，数学是一门关于鉴别和选择的学问。那么，期刊收录数学论文的标准是什么呢？为什么费马大定理的证明会被认为是20世纪最伟大的数学证明之一，而同等复杂程度的数值计算却是平庸而无趣的？“当 $n > 2$ 时，方程 $x^n + y^n = z^n$ 没有整数解”的证明到底有趣在哪里？

这就是为什么我说数学不仅是一门有用的科学，而更像是一门创造性的艺术。定理证明的叙述，是决定这个定理能否在数学的万神殿中占据一席之地的重要因素。因此，我相信一个好的证明就像一个动人的故事，抑或是一首美妙的乐曲，可以启发或引导“听众”踏上转变之旅。

## 数学寓言

通过讲故事的方式可以更好地解释数学证明的叙述质量这一概念。我13岁时读了哈代的《一个数学家的辩白》，这是我第一次接触数学证明。该书描写的是一名数学家的切身感受。格雷厄姆·格林（Graham Greene）认为该书对创作型艺术家的描述是继亨利·詹姆斯（Henry James）的日记以来最贴切的。

书中提到了欧几里得发现的，极可能是数学史上最早的一个证明。如果把这个证明看作一个故事，那么故事的“主角”就是素数。素数又称质数，是一个大于1的自然数，且除了1和它本身外，不能被其他自然数整除。例如，3、7、13，等等。现在我们一起踏上叙事之旅，揭开关于故事主角的谜题——素数是无穷无尽的这一特性。本章开始部分已经介绍了Mizar系统对该定理的证明。现在，由我来告诉你这个故事。

证明就像数学家的旅行游记，欧几里得通过他心灵的窗户看到了这样的景象：素数就像一座座山峰，重峦叠嶂，绵延不绝。后辈数学家们肩负的任务就是寻找一条从熟知的领域出发，通向这片未知新世界的道路。

就像《指环王》中弗罗多从夏尔到魔多的冒险一样，证明就是对这段旅程的描述。在夏尔这片人们熟悉的土地上有数学公理（关于数学的不证自明的真理）以及那些已被证明的命题，这是任务的初始设置。从故土出发的旅程受到数学推导规则的限制，就像棋类游戏的行棋规则，这些规则确定了通过这个世界的行进路线。偶尔陷入僵局后，你需要绕道而行、侧路包抄，甚至以退为进。有时候，你需要等待新角色（如虚数、微积分）的加入才能继续前进。

证明是一场“按图索骥”的旅程，地图上标定了穿越的路径。成功的证明是一组路标，指引所有后辈数学家走完相同的旅程。证明的读者们将通过地图所指的道路抵达遥不可及的高峰，体会到和作者一样的惊喜和感动。很多时候，证明不是寻找 $i$ 和 $t$ 的交点，就像故事不会呈现某角色的每个生活细节——它是对整个旅程的描述，而不是具体步骤的重现。数学家提供的论据旨在引导读者的思想。哈代将论据描述为：“为打动某些人而编造的一堆华丽辞藻；讲演时用来演示的图片；激发小学生想象力的工具。”

结尾即是故事的开始，倒叙是数学故事最特别的地方。问题在于故事情节如何设计才能从当前背景到达这一高潮。叙事之旅需要进行一些场景设置——简要的前情描述，告诉我们素数的重要特征是它们是所有其他数字的约数，即每个数都可以由一个或者多个素数相乘得到，例如  $105=3\times5\times7$ ， $16=2\times2\times2\times2$ 。

因此，让我们开启旅程来解释为什么素数有无穷多个。用反证的方法，假设素数不是无穷的，我们可以一一列出这些剧中的“角色”。反证法是数学家工具箱中常用的叙事工具，就像《爱丽丝梦游仙境》或《绿野仙踪》一样，想象出一个完全相反的世界，并试图证明这个世界是真实的，直到故事以一个荒谬的结局告终。最终的结论说明先前的假设是错误的。

我们假设剧中人物（素数）由2、3、5、7、11、13组成。不难看出，有人被漏掉了（例如，17是素数，但不属于剧中的人物）。将字符相乘：

$$2\times3\times5\times7\times11\times13$$

然后，将得到的结果加1：

$$2\times3\times5\times7\times11\times13+1$$

这一步就像是短篇小说故事情节中出现的一个转折点，会将剧情导向一个完全出乎意料的结局。

这个新数字必须通过剧中人物（已有的素数，我们开启旅程时熟悉的外部环境）来构造。那么，哪个素数可以整除这个新数呢？它不可能再是剧中的人物了，因为有一个余数1。但由于素数是所有其他数字的约数，所以一定存在一些素数可以整除这个数。这意味着我们在设定剧中人物时，漏掉了这些素数。实际上，这个新数可以通过素数59和509相乘得到。

你或许会建议我将这些新角色添加到剧中，但这个故事的有趣之处就在于，它可以再讲一遍，可结果是你会发现仍然缺少一个角色。依此类推，任何有限的素数列表都会丢失一些素数，因此，素数的个数必须是无穷的。

证明完毕！ [1]

[1] 数学家喜欢在证明的结尾写一个QED的标记，其源自拉丁语quod erat demonstrandum（意为“这被证明了”）的缩写。——译者注



## 意料之外的故事

对我来说，数学证明最重要的不是追求“证明完毕”，也不是得到的最终结果，而是整个证明的过程，即通向目的地的旅程，这就像音乐的全部并不是最后的一个和弦一样。知道素数有无穷多个这一结果固然很重要，但我们的满足感源自对其原因的理解。在阅读小说和证明数学定理时，最令人开心的莫过于将所有的线索聚集在一起揭开谜团的那个顿悟的时刻。比如，将某段音乐中的和声结构分辨出来，或者将某个谋杀之谜解开。

“令人惊讶”是数学的重要特质。数学家迈克尔·阿蒂亚（Michael Atiyah）这样描述他最喜爱的数学特质：

我喜欢惊喜。规行矩步的论证是枯燥无味的。我喜欢意料之外的收获：全新的视角；与其他领域的交叉融合；在故事尾声处发生的大逆转。当在创作一篇新的数学作品时，我所做的选择将被一种愿望所激励，那就是带领我的读者踏上一段充满曲折和惊喜的有趣旅程。我想用一个问题跟我的读者开个玩笑：为什么两个看似毫不相干的角色应该建立关联？然后，随着证据的展开，人们会逐渐意识到，或者说突然意识到，这两个角色（概念）实际上是一样的。

我最喜欢的定理之一是费马发现的关于某些类型的素数具有的一个奇特性：如果一个素数除以4后所得余数为1，那么该素数等于某两个数字的平方和。例如，41是素数，其除以4后余数为1，而41又可以写成 $25+16$ ，即 $5^2+4^2$ 。但是，这适用于所有类型的素数吗？除以4后余数为1的素数有无穷多个，为什么它们会与平方数有关？

对于这个故事开头，我最初的反应是难以置信。但当费马带我踏上证明之旅后，我看到了素数与平方这两个不相关的概念建立联系、融为一体，获得了巨大的满足感。它们就像一段复调音乐，两个声部以不同的主题同时进行，但最终仍然以非常和谐的方式融为一体。

这里举一个更简单的例子，我在第9章中提到的小游戏：将连续的奇数相加，结果会如何？

$$1+3=4, 1+3+5=9, 1+3+5+7=16, 1+3+5+7+9=25$$

N个连续的奇数相加，其结果等于第N个数的平方，其证明如图13-1所示。

这种满足感来自从奇数到平方数的意外之旅。当突然明白了为什么这两个明显不相干的角色之间会有联系时，我体会到了豁然开朗的喜悦。

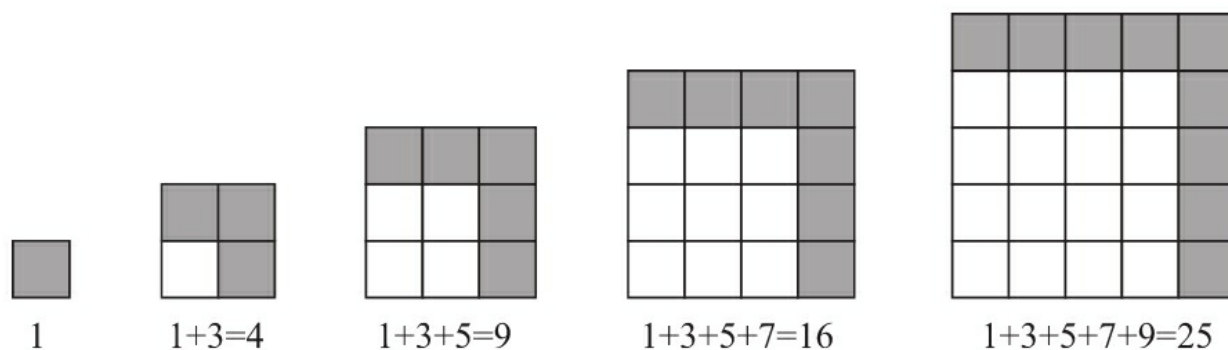


图 13-1

寻找椭圆曲线的解是数学领域最棘手的问题之一。我构建出一种新的对称元，并发现了其子群结构与椭圆曲线模p解的数目计算的相关性。在学术研讨会上给数学同行提供的证明以及在期刊上发表的论文中，我都详细陈述了数学世界的这两个截然不同的领域是如何关联的。我对自己的“数学故事”津津乐道，原因在于我喜欢看到在我的启发之下，倾听者突然顿悟时呈现在脸上的那种喜悦之情。数学家的艺术不只是创造出新的东西，还包括讲述一个令人惊讶的故事。正如庞加莱所言，它是一种选择。

就像有时候人们在读完一本伟大的小说之后会感到悲伤，数学家在探索终结时也有自己的悲伤。我们一直享受着费马大定理带来的证明探索之旅，所以当安德鲁·怀尔斯揭开这个有着350年历史的谜题时，我们有着既快乐又悲伤的复杂之情。这就是为新故事开辟道路的证明会如此重要的原因。

## 数学的叙述艺术

“悬念”这一特性是数学证明故事中经典的叙事工具。故事伊始，作者使用情节元素提出问题，让读者带着解决问题的目的继续阅读。这种叙事方法被称为阐释代码<sup>[1]</sup>，是罗兰·巴特（Roland Barthes）提出的五种关键叙事代码之一。它是未解之谜（或未答之题）给出令人满意的数学证明的核心方法。当我们研究数学时，能给我们带来愉悦的就是那种想要解开谜团的渴望。从这个意义上说，数学证明与一部精彩的侦探小说有很多共同之处。

数学证明都是从故事的结局开始。科幻动作或谋杀悬疑题材的作品也有类似的剧情设置。比如，《星际迷航：下一代》就从整个故事的结尾处开始：企业号星舰陷入一片火海，皮卡德下令弃船，紧接着飞船就爆炸了。虽然大多数文学作品并不追求如此戏剧性的开场，但其内部也经常会有这种时空颠倒的叙事情节。

除了开场环节通过未解的问题制造的紧张感之外，数学故事的另一个叙事驱动力源自证明展开时的内在行动，它是通过故事情节的延续推动叙事逻辑沿着时间轴向前发展的动力。欧几里得的证明所涉及的问题是“素数有无穷多个”，证明中把素数作乘法运算后得到一个新数字。读者在心中产生疑问的同时被激发了进一步了解的兴趣：“为什么要这么做？这个新数字将作何用？”此时，行动建立。将新数字加1后，读者会更加好奇。这一系列行动结束后，故事情节将发展到高潮，最终获得解决问题的启示和方案时，读者将收获极大的满足感。这就是巴特总结的五种叙事代码中的第二个——行动代码的一个很好的例子。一系列动作的累积制造出悬念，而动作本身又隐含了下一步的叙事动作。

巴特的另外三个代码是语义代码、符号代码和文化代码。这三种代码均围绕一个设计意图展开，即故事中的某些思想会与故事之外的事物产生共鸣，从而赋予其更多的意义。这三者都是构建数学证明的重要工具，发掘读者已有的知识以获得证明的预期效果。就像哈代所说的那样，有时候证明需要在大量历史知识或观点的“触发”下推进。就像文学故事一样，如果利用不好这些触发条件，就会大幅降低证明的效率。

故事的总体叙事也被称为故事的原型或者主线。文学理论家们把各种故事原型进行归纳和总结，最终确定了七种不同的叙事类型，比如灰

姑娘型故事、探险型故事、战争型故事等。数学故事有没有主线？当然有。数学家识别出某些证明原型，并引用其方法来帮助读者。证明方法有反证法、归纳法、概率分析法，等等。费马大定理的证明就采用了反证法：假定原命题存在对立面，按照推理规则进行推演后发现它是正确的，那么原命题就是错误的。怀尔斯的证明从假设费马方程有解开始进行探索，最终得到有悖常理的结论，而这一结论意味着初始假设是错误的。

好的数学有一种张力<sup>[2]</sup>，其证明既不会很复杂也不会很简单。完美的证明有其必然性，但每一步都无法提前预测。约翰·卡维尔蒂（John Cawelti）在他的《冒险、神秘和浪漫》（Adventure, Mystery, and Romance）一书中对文学作品张力的描写同样也适用于数学：“追求秩序和安全的结果可能导致单调乏味和千篇一律，但为了创新和改变而不顾秩序，则会带来危险和不确定性……文化的历史可以被诠释为在追求秩序和避免乏味之间的动态张力。”

这种追求是一个好的证明的核心。

很少有专业的数学家听说过Mizar项目，因为它的目的就不是让人真正感兴趣。Mizar构建的是看似包罗万象实则一无所有的“巴别图书馆”。难道它不能从我们喜爱的数学中学习，去创造我们喜爱的数学吗？当然能，它只是科技在发展过程中的一个“短暂停留”，我仍然相信机器学习的潜力。

尽管大多数人认为音乐是与数学相关的创造性艺术，但在我看来，讲故事是最接近证明定理的创造性行为。如果数学证明是故事，那么计算机在讲故事方面的能力会有多强呢？

[1] 阐释代码也称为“谜的代码”，指的是类似于侦探小说中具有设谜和解谜功能的句段。只要文本中有需要揭示的真相、需要澄清的谜团，那么这个文本就含有阐释代码，所以巴特称之为“真相的声音”。——译者注

[2] 张力本意是让水滴圆润凝聚而不分散的力量。若某首诗具有张力，说明这首诗全篇对中心观点的凝聚感十分强烈。——译者注

## 第14章 语言游戏

### Twitter笑话一则

两位科学家走进一家酒吧。

一位说：“我要一杯一氧化二氢。”（I'll have H<sub>2</sub> O.）

另一位说：“我也要一杯一氧化二氢。”（I'll have H<sub>2</sub> O, too.）

然后，调酒师给了他们两杯水。这是因为调酒师能够辨别在句末指示不同语法功能的同音异义词（即调酒师进行了准确的判断，给了他们两杯水，而不是一杯水、一杯双氧水）。

如果你想成为一名作家，理解语言是很重要的，或者至少要有理解语言的愿望。那么，机器在此方面有多擅长呢？艾伦·图灵在其著名的论文——《计算机器与智能》的开篇提出了这样的问题：“我提议考虑这样一个问题：机器会思考吗？”后来图灵认为这样问太泛泛了，所以他改进了这个问题：他想知道是否可以给一台机器编程，然后让它与人类交流，看机器的反应能否令人相信跟其交流的是“人”而不是一台机器。

图灵称此为“模仿游戏”（这是根据当时流行的一种游戏命名的），就是现在大名鼎鼎的“图灵测试”。若要通过图灵测试，算法需要能够接受千变万化的“自然语言”作为输入，并对其进行处理，然后生成与人类可能做出的反应相对应的输出。（“自然语言”一般指人类在自然进化过程中不断被重复、被使用的，没有经过有意识的预先设想和计划的语言。其与计算机代码有着鲜明的区别。）

第一个成功通过图灵测试的算法，是计算机科学家约瑟夫·魏泽鲍姆（Joseph Weizenbaum）在1966年开发的名为“ELIZA”的程序。魏泽鲍姆的想法既聪明又极其简单。如果你曾经看过心理医生，而这名心理医生刚好接受过卡尔·罗杰斯<sup>[1]</sup>（Carl Rogers）的以当事人为中心的治疗培训，你就会了解，心理医生会问一些问题，但不会把谈话导向一个新的方向。他问的这些话仅仅是继续了前面的话题，而这为患者提供了进一步阐述其精神状态的空间。



患者：男人都是一样的。

心理医生：在哪一点上？

患者：他们总是在这件或那件事情上令人烦躁。

心理医生：比如呢？举个例子。

患者：比如到这里来，就是我男朋友让我来的。

心理医生：是你男朋友让你来的？

患者：他说我大部分时间都很抑郁。

心理医生：那你认为你大部分时间都很抑郁吗？

在上例中，心理医生不是一个人，而是一个电脑程序。魏泽鲍姆可以通过算法捕捉罗杰斯式心理医生在治疗过程中对谈话问答的反应。他设置了一套规则，用来接受患者的输入——“如果.....那么.....”，然后通过分析句子的结构，输出一些基于输入内容的问句。例如，患者说“I am X”（我这样了），ELIZA会输出“How long have you been X？”（你这样多久了？）。如果X=sick<sup>[2]</sup>，那么ELIZA会回答“你生病多久了？”这似乎是合适的。如果X=sick of you（讨厌你），那么按照前面的规则，ELIZA的回应就不太有效了。因此，ELIZA遵循的另一个规则是将所有的第二人称单数代词改为第一人称单数代词，将第一人称单数代词改为第二人称单数代词。那么，现在其回答就变成了：“How long have you been sick of me？”（你讨厌我有多久了？）

其中的关键是要编写很多的规则来让对话令人信服。如果输入与ELIZA设定的场景不匹配，那么它会聪明地邀请患者继续说下去：“再多告诉我一些。”

与ELIZA的互动是如此的真实，以至于有报道称，魏泽鲍姆的秘书曾要求魏泽鲍姆回避一下，以便她可以和ELIZA单独交谈。还有一个例子，魏泽鲍姆让一位来访者前往实验室，在不知情的情况下配合进行试验。这位来访者在电脑键盘上发现了一张纸条，上面写着：“请在这台电脑上键入信息，并以句号结尾。”结果，他并没有意识到自己是在和ELIZA进行交流，自始至终，他都以为自己在和实验室的操作员进行远



程交流。交流内容如下：

来访者：请问今天早上我可以使用电脑吗？

ELIZA：你为什么这么问？

来访者：也许我可以做一些额外的网络销售。

ELIZA：你为什么不太肯定？

来访者：我觉得是这样的……客户从来没体验过该系统的演示。

ELIZA：这对你很重要吗？

来访者：当然，很重要。

ELIZA：你似乎对这事很是乐观。

来访者：太荒谬了。

ELIZA：怎么了？什么太荒谬了？

来访者：为什么我现在不能使用电脑？

ELIZA：再跟我说说你的看法……客户。

来访者被这些含糊其辞的回答激怒了，他要求和操作员直接对话。他气愤地输入了如下内容：

来访者：请回电491-1850

因为他忘了用句号来结束句子，所以ELIZA没有回答他。他非常生气，拿起电话直接拨给了操作员，他以为他拨给了那个一直和他说话的人。他问道：“你为什么如此怠慢我？！”接到电话的实验室工作人员一头雾水。当来访者把对话复述给他听时，实验室工作人员哈哈大笑起来。来访者费了一番功夫才弄明白刚刚和自己交谈的是一台电脑。

也就是说，与ELIZA交谈没多久，人们就能察觉出它的回答有些地方不太对劲：应答的范围有限也不灵活，并且对以往的对话没有记忆、

联想。之所以其治疗情景中应用效果良好，是因为治疗通常取决于患者是否能解决他们自己的问题。作为医生的算法应能诊断病情，所以ELIZA并不擅长进行疾病诊断，但另一些算法可能会比你的全科医生更擅长疾病诊断。如果你对此很感兴趣，可以通过互联网上许多的在线版本对ELIZA进行测试。

20世纪90年代初，发明家休·勒布纳（Hugh Loebner）赞助设立了勒布纳人工智能奖<sup>[3]</sup>（Loebner Prize）。该奖项旨在遴选在人机交互方面更具灵活性和说服力的程序。勒布纳是一名发明家，但是人工智能领域的门外汉。他邀请了一个评委小组在一段时间内分别与人类和人工智能进行互动，并判断哪一个是人工智能。如果一个程序能够获得半数评委的确认票，那么它的创建者将会获得一枚银质奖牌和2.5万美元的奖金。如果没有任何一款程序达到银奖的标准，那么奖金将归属于能使最多评委相信“它”是“人”的程序的创建者。

第一个通过扩展版图灵测试的程序将获得一枚金质奖牌和10万美元奖金。这个测试的内容包括破译和理解文本、视听组件等内容，通过标准是评委们无法将其与真正的人类区分开来。这样的程序出现意味着什么，目前还没有很明确的说明，但可能意味着你看到和听到的是机器人，但你仍然会确信它是人类。一旦此奖项被颁发出去，勒布纳人工智能奖就终止了。勒布纳乐于这样宣传：他所提供的金质奖牌是纯金的，不像奥运会奖牌那样是镀金的。

到目前为止，还没有任何一个程序获得了银牌，就更不用提金奖了。近来，在此赛事中取得较好成绩之一的是一个名为“三井”（Mitsuku）的聊天机器人程序，化名为米莉（Millie）——一位利兹大学的学生。它的作者是史蒂夫·沃斯维克（Steve Worswick）。他原本运营着一个音乐网站，后来给网站加了一个聊天机器人，结果他发现浏览者花了大量的时间与机器人聊天，于是他决定将所有精力集中在开发聊天机器人上。

在2013年<sup>[4]</sup>获得第23届勒布纳人工智能奖银奖的“三井”，在与人们对话的时候，会对人们提出的问题给出一些非常有见地的回答：

问：你对特朗普有什么看法？

“三井”：据我所知，有超过50%的人担心特朗普会当选总统。如果真是这样的话，那么他们为什么要投票给他呢？

问：你还有什么别的看法吗？

“三井”：目前为止这就是我所知。

然而，很快“三井”对其他问题的应答就显示出了奇怪的运算失误：

问：34 957加70 764等于多少？

“三井”：34 957加70 764，嗯，对的，34 957是一种用羊奶制成的奶酪，但它并不怎么受人们欢迎。

在此之前，人与“三井”之间的对话交流都很顺畅，这让人对人工智能充满了信心和希望。但是，它现在却以这样奇怪的方式戛然而止了。如果你曾经和聊天机器人聊过天，那么你就用你跟它的对话有效地训练过它。其中一些聊天机器人的算法是通过存储人类对特定问题的反应来工作的，因此，如果聊天机器人在未来遇到类似的情况，它可以逆转角色，使自己应对起来更像是人类。但大多数算法都是基于ELIZA“如果.....那么.....”的规则和更为复杂的规则，因而，这些算法永远无法真正处理语言的多样性。我们需要理解语言是如何组合在一起的。

人工智能系统在尝试处理“自然语言”时，会在语言模糊性和语言情境方面存在一些问题。在勒布纳测试中通常会进行一组名为“威诺格拉德”的测试（该测试由斯坦福大学教授威诺格拉德提出），在这个测试中那些不能很好理解语言模糊性和语言情境方面问题的人工智能很快就会原形毕露。如下句中的“they”（他们）：

The city councilmen refused the demonstrators a permit because they[feared/advocated]violence.

市议员们拒绝签署对示威者的许可，因为他们（害怕/主张）暴力。

此句填空的选择，很明显会根据“they”的指代不同而变化。基于对语境和过往经验的认知，人类知晓如何理解这个谜语似的句子，但是对于机器真的很难。“威诺格拉德”测试正是利用了“自然语言”的复杂性、丰富性和模糊性等特征。

“三井”在2017年的图灵测试中就遇到了这样的“威诺格拉德”测试<sup>[5]</sup>

:

I was trying to open the lock with the key, but someone had filled the keyhole with chewing gum, and I couldn't get it out. What couldn't I get out?

我要用钥匙打开锁头，但有人用口香糖把钥匙孔填满了，我弄不出来它。阻止我弄出来的是什么？

The trophy doesn't fit into the brown suitcase because it's too small. What is too small?

奖杯不适合放进棕色的手提箱里，因为它太小了。什么太小了？

我们该如何培养驾驭语言复杂性的能力？人类所使用的代码是长久以来在与他人的语言交流中逐步形成的。从幼年开始，人类就在语言环境中浸染，在这个过程中我们不断地学习，不断地犯错，不断地再学习。有了机器学习的新工具，算法最终能学会怎样处理“自然语言”吗？互联网上有大量的可作为数据库使用的实例。所以，我们为什么不能将算法投入互联网任其自由发展，去学习驾驭语言的复杂性、模糊性呢？

语言学家们惊奇地发现，一个孩子想要理解其他人并与之互动，所需要的语言真的不用很多。诺姆·乔姆斯基（Noam Chomsky）认为语言天赋是我们与生俱来的，就好像我们是按照老式的自上而下的模型进行编程的，而不是从零开始学习。如果真的是这样，对机器学习来说，仅通对一个庞大的语言使用数据库来学习语言将会是非常困难的。

[1] 1902—1987，美国心理学家，人本主义心理学的主要代表人物之一。他从事心理咨询和治疗的实践与研究，主张“以当事人为中心”的心理治疗方法，首创非指导性治疗（案主中心治疗），强调人具备自我调整以恢复心理健康的能力。——译者注

[2] sick的意思有厌恶的、病态的、不舒服的、渴望的、生病的、恶心的，这里应取生病的意思。——译者注

[3] 勒布纳人工智能奖分设金、银、铜奖。金奖、银奖的奖金分别为10万美元、2.5万美元。铜奖又分为四个等级，奖金分别为4000美元、1000美元、500美元、200美元。如果程序不仅能以文本方式通过交谈测试，在音频和视频方式的测试中也能过关，则获金奖，赢得10万美元和一枚18K金质金牌；如果它能在比较长的时间内以文本谈话的方式迷惑

至少半数评委，则获银奖；在未达到以上标准的程序中，迷惑住最多评委的将获铜奖。——译者注

[4] 原文中记载的时间为2017年，经译者查阅资料并综合周边信息分析，“三井”获得第23届勒布纳人工智能奖银奖的时间应为2013年9月14日，特此勘误（<https://www.pandorabots.com/botmaster/en/home>）。——译者注

[5] 因为语种不同，思维逻辑不甚相同，所以建议读者看英文体会。——译者注

## 《危险边缘》

十几年前IBM的超级计算机“深蓝”战胜了国际象棋冠军加里·卡斯帕罗夫（Garry Kasparov），从他手中夺取了国际象棋的桂冠。2011年，IBM将注意力转向了与国际象棋或围棋截然不同的一个方向：挑战变幻莫测的“自然语言”——参加益智问答节目《危险边缘》<sup>[1]</sup>。

《危险边缘》是一个常识问答节目。考虑到对计算机来说上网搜索维基百科是相当简单的事，这看起来不像是一个对于算法的测试。那么难点在哪里呢？更多的困难来自问答的方式。节目以一种问答倒置的方式进行，以答案的形式提问、提问的形式作答。例如，题目是“这是一个元素，在元素周期表上的序号是27，它可以制造‘绿’和‘蓝’的颜料”，答案是“是‘钴’吗？”（What is ‘cobalt’？）

决胜《危险边缘》！这个过程包括理解、解析问题，访问大型的知识数据库，并尽快选择可能性最高的答案。《危险边缘》常常利用双关语、文字游戏、转移注意力等方法来迷惑参与者，即使是人有时也很难理解其问题的意思。由这些因素带来的模糊性使得算法几乎不可能100%准确。但对于IBM来说，它并不需要做到100%准确，只需要比其他选手好就足够了。尽管在IBM内部仍有一些人认为，把时间花在这样一个微不足道的问答游戏节目上是在浪费资源，但其他人坚持认为，赢得比赛将标志着机器在解析语言语义方面的能力有了重大提升。

如果卡斯帕罗夫是国际象棋之王。那布拉德·拉特（Brad Rutter）和肯·詹宁斯（Ken Jennings）就是《危险边缘》之王，他们两人都创造了非凡的战绩。詹宁斯已经连胜74场，而拉特通过该节目赢取的奖金已超过400万美元。虽然拉特一直被认为是一个“学痞”，但两人在学校的智力测验小组中都曾崭露头角。《危险边缘》中通常有三个参赛者，这两个人类的“王者”同意和IBM的算法“华生（Watson）”同台竞技。IBM的算法“华生”，不是以夏洛克·福尔摩斯的助手名字命名的，而是为了纪念该公司的第一任首席执行官托马斯·J·沃森<sup>[2]</sup>（Thomas J. Watson）。

2011年1月，拉特和詹宁斯在两天多的时间里勇敢地与“华生”决战。由于无法将电脑硬件设备搬到电视演播室，节目的拍摄工作不得不在位于纽约州约克城高地的IBM研究实验室进行。除了地点不一样外，节目一切如常，主持人亚历克斯·特里贝克（Alex Trebek）像往常一样



主持节目并发问。节目在全美电视网上播出，让所有人看到了人类被机器超越。

开局阶段，人类选手发挥得不错，并在一段时间内保持了领先，但最终还是不敌IBM算法的威力。要赢得比赛不仅仅要擅长回答问题，有时也需要一定的策略：参赛者可以在最后一个题目上下赌注，这使得落后的选手有可能在最后一题时将手里的钱数翻倍，从而赢得比赛。正因为如此，“华生”也分出了一部分算力，来确保它能有效地运用所有数学技能进行下注。

“华生”在比赛中似乎有一个不公平的优势：抢答器。按照规则，只有先按下抢答器的选手才有资格回答问题。一开始“华生”被允许使用电子信号的方式按抢答器，而不是像人类那样按下实物按钮。但人们很快意识到，这将给“华生”带来巨大的优势。因此，在发现这个漏洞之后，“华生”就需要激活一个机械手指来按下按钮了。虽然这样使“华生”的速度稍慢了一点，但它在这方面仍然比人类快得多。就像詹宁斯说的：“如果你想赢，那么抢答器就是一切。”问题是“华生”每次都能在微秒级精确地按下抢答器。在这方面，人类的反应能力和计算机电路相比差远了。其实，这其中也有一定程度的运气因素：答题的过程中在题板上会给出所谓的“双赌法”<sup>[3]</sup>，“华生”很幸运地在比赛中选到了。如果人类选手运气好的话，究竟鹿死谁手还不得而知呢。

尽管“华生”赢得了比赛，但它确实犯了一些很明显的、常识性的错误。例如，比赛中有一题是猜美国的一座城市，题面是：该城市最大的机场是以一位二战英雄的名字命名；该城市第二大机场以二战的一场战役命名。人类做出的正确回答是“芝加哥”，而“华生”却回答“多伦多”。多伦多是加拿大的一座城市，连美国的边都不挨呀！

负责研发“华生”的IBM研究员大卫·费鲁奇（David Ferrucci）说：“对于当时所发生的我们也了解得不够。现实状况是这样，有很多数据的标题是美国的城市，但内容确实是其他国家、欧洲城市、人物、官员……也就是说，尽管‘华生’认为某名称是美国的城市，但我们对其判别的数据特征是存疑的。”值得称道的是，“华生”也对答案不那么确定，它在自己的答案之后加上了一串问号。这个问题是需要下注的，在下注的环节，“华生”依然对自己的答案表示不那么确定，所以它下了很小的赌注。

在最后一个问题提出时，当时“华生”很明显已取得了胜利，故詹宁

斯在回答“布拉姆·斯托克”后紧接着补充道：“欢迎我们的新电脑霸主。”这是在模仿《辛普森一家》中的桥段，该桥段是对1977年的一部B级电影——《蚂蚁帝国》（改编自H.G.威尔斯（H.G.Wells）的作品）的恶搞（该电影中的一个角色以这种方式屈服于一只巨大的昆虫）。

显然，“华生”并没弄明白这种幽默。

[1] 哥伦比亚广播公司的益智问答游戏节目，已有数十年历史。该节目以一种独特的问答形式进行，问题涵盖的知识面非常广，涉及历史、文学、艺术、流行文化、科技、体育、地理、文字游戏等领域。与一般的问答节目相反，《危险边缘》以答案的形式提问、提问的形式作答。参赛者需具备广泛的文化知识，还得解析隐晦含义、反讽与谜语等，而电脑并不擅长进行这类复杂的思考。——译者注

[2] “Watson”曾被译为“华生”。——译者注

[3] 通常指在赛马博彩中，选择两场次指定的比赛都胜出即为赢的押注法。——译者注

## “华生”的工作模式

要解释“华生”的工作原理，最好的方法就是请你想象一个巨大的场景，其中有单词、名字和其他可能的答案，它们散布在各处。IBM所做的第一步是以某种连贯的方式排列单词；第二步是理解每个问题，并为该问题生成候选位置标记。

这个场景不同于你从窗口望出去所看到的三维世界的样子，而是一个复杂的数学场景：不同的特定属性由不同的维度衡量，这些特定的属性将决定一个单词所拥有的特性和品质。识别和选择这些特性和品质是一门艺术，例如某个词可能与地理或年代具有很强的关联，抑或与艺术或体育相关。当然，它也可能与几个不同的性质相关，在这种情况下，它的位置将被定位在这几个相关的属性之下。例如，爱因斯坦会演奏小提琴，那么他就会被定位于“科学家”和“音乐家”这两个方向。但你不会从音乐家的角度继续推演，而是会从科学家的角度继续推演。IBM团队分析了2万个样本问题，发现了大约2500种不同的答案类型，其中大约200种类型涵盖了被问到的50%以上的问题所对应的答案。

“华生”算法的运行要经历四个阶段的分析：

（1）分析问题，以确定可能的答案范围。

（2）在候选范围里根据问题选择大约200个可能的答案。

（3）评价这些可能的答案。评价过程是将这200个多维点压缩在一个数值维度上，然后对这些答案进行评分并依次排列，最后对答案的信任值进行评价。

（4）给出答案。当评分以及信任值评价高于一个设定的阈值时，算法就会给出建议的答案。

以上过程要在短短的几秒钟之内完成，不然人类选手将会抢先按下抢答器。

让我们考量这样的问题：

黑洞<sup>[1]</sup>（THE HOLE TRUTH）：1756年6月20日夜，亚洲地区，众

所周知的不好的事件。

答案是：加尔各答黑洞事件。在这一题里，该答案的时间和地点的评分很高，大概同一时间、同一区域也发生了其他的不好的事件。答案中“hole”这个单词，更加有助于“华生”的评价。到这一步时，“加尔各答黑洞事件”的评分和信任值评价就远高于在同一时间、同一区域内发生的其他事件了。所以“华生”给出了正确的答案。

当出现“写作”“作曲”“笔”“出版”这样的词语时，算法将推动你走向艺术创作的范围。“普希金写成的第一首诗”这个线索会把我们引向“作者”所在的答案区域。一旦选定了200个候选选项，算法马上会开始对这些候选选项进行评分，这个过程就需要仔细权衡所选取的每个答案的不同维度的权重。算法必须要找出一种方法来衡量假定的答案与问题的接近程度。能与维基百科中的内容进行准确的语义匹配的数据，在其他限制因素也匹配的情况下，很可能会得到非常高的分数。例如，“人，男性，1594年，安达卢西亚，做税务员”。关于这个问题，答案“梭罗”和“塞万提斯”在语义匹配上的得分都很高。但是，“塞万提斯”生卒年信息是“1547—1616”，而“梭罗”生于1817年，故“塞万提斯”比“梭罗”在时间维度上更匹配，所以在时间维度上“塞万提斯”得分更高。

“华生”研究团队提出了50种不同的评分模型。在未知评分情况的状态下，“华生”会广泛地选择候选答案：采用广泛撒网、重点培养的模式运作，即先倾向于选择包含多种可能的答案，然后利用评分机制选取得分靠前的几个选项。这有点像选择旅店入住。首先你需要选择想入住的区域，并搜集该区域所有旅店的信息。这时你会根据网上的价格和评价信息进行选择，但这样的选择方式就很有可能有利于一家值得入住但位置偏远的旅店。

该算法进行评分的方式允许它以自下而上的方式从错误中学习并细化参数，有点像转动刻度盘来重新定义函数。这其中的艺术在于算法试图为你拨动的刻度盘找到最佳的设置，在尽可能多的不同的背景下得到正确的答案。让我们考量一下这样一个问题：“国家，与智利的陆地国界最长。”与智利接壤的国家有两个，分别是阿根廷和玻利维亚。对于这两个候选答案，你如何进行评分呢？如果其中一个经常出现在人们的视野里，那么人们就有可能给它更高的分数。在这种情况下，玻利维亚可能会得到更高的分数，因为智利和玻利维亚在边界问题上有很多纷争，新闻报道上进行着长篇累牍的报道。但如果使用更具有地理性质的原始资料数据，并对这两个国家在出版物中被提及的次数进行评分，那

么阿根廷的得分就会高很多，因为它就是正确答案。

当詹宁斯得知“华生”是如何运行的时候，他大吃一惊：“电脑破解《危险边缘》的方法，听起来和我的差不多。”詹宁斯会在线索中找到关键词，然后在他的记忆中搜索，并找出与这些词相关联的一串串词（“华生”可以访问一个容量为15TB的人类知识数据库）。然后，他根据自己收集到的所有信息将它们进行关联性的考量，仔细地考虑最可能的选项。关联性的线索包括：线索中暗示的时间、地点、性别，抑或是体育、文学、政治等范畴。詹宁斯说：“对于人类选手来说，这是一个即时的、直观的过程，但我确信在这背后，我的大脑或多或少在做着同样的事情。”

为什么IBM要付出这么大的人力、物力、财力去赢得这样一个问答游戏比赛？这么做看起来似乎没有什么实际意义，但对于IBM和DeepMind这样的公司来说，赢得比赛为其提供了一个相当明确的成功指标。比赛要么赢要么输，没有模棱两可的暧昧。赢得比赛这件事给需要销售产品的公司提供了极好的宣传噱头，因为人人都喜欢“人机大战”的戏码。这就像一场算法的“时装秀”，让计算机公司展示自己傲人的编码能力。

IBM的“华生”已经改变了我们对计算机的认知——它不仅击败了所有选手问鼎《危险边缘》的冠军，还被应用于医学诊断领域。是什么让“华生”与众不同？它又有什么不同？通过我们对它的训练，“华生”对非结构化数据进行处理的能力是它最大的优势所在。此外，人们只需将文本输入“华生”，实际上就组建了一个可靠的可以理解文本内重要信息的系统。“华生”在参加《危险边缘》之前，就已经将整个维基百科进行了离线存储，以供其随时调用。此外，人类可以告知“华生”哪一个信息来源更为可靠。IBM将这种成果称为“认知计算处理”，这是因为人类在此过程中的角色从调度安排转变成了训练。

在未来，我们将减少死记硬背的计算，更多地依靠互动和学习。算法非常聪明，只要有更多的信息，它就能够做到排除某些答案，或者提高对某些已经提供的待选答案的信心值。当“华生”在当前的应用程序中处理一个困难的问题时，它可以生成一组可能的答案集，也可以使提出的问题趋于简单明了。大多数这种问答系统都是为处理一类特定的、边界明确的问题而设计的，这就意味着它只能回答某些类型的问题，而且在输入问题时必须要以特定的方式进行措辞，方能获得算法的响应。“华生”可以处理“开放域”的问答，这意味着面对“华生”，你想问什

么就可以问什么。它使用“自然语言”处理技术来分解你抛给它的词句，这样可以使算法真正理解问题，即便你在问它时使用了不寻常的方式，它也能有所回应。

IBM发布过一个关于“华生”和IBM DeepQA（DeepQA是“华生”在生成假设时使用的一项基础技术）的FAQ（常见问题解答），这个常见问题解答非常详尽、有用。《星际迷航》中的电脑就是一个合适的范例。这个虚拟的计算机系统可以被看作一个交互式的系统，它可以回答任何问题，还可以提供关于任何主题的精确信息。

[1] 加尔各答黑洞事件：加尔各答黑洞是一座用来监禁英国俘虏的场所，据说是一间环境极为恶劣的普通小土牢。1756年6月20日，被俘的146名英国人中有123人在此处因窒息死亡。此事件引起了国际争论，亦为英法两国为争夺印度半岛殖民利益所发生的纠纷战争中著名的历史事件。很多学者对此事件（即“黑洞悲剧”）进行过论述，认为未必可信。  
——译者注



## 在翻译中迷失

虽说在学校里学习语言的时候很痛苦，但我还记得在《银河系漫游指南》（The Hitchhiker's Guide to the Galaxy）里看到过的“巴别鱼”（babel fish）的故事。“巴别鱼”是一种黄色的、类似水蛭的小动物，当它掉到你的耳朵里时，会以脑电波为食，即时翻译出你听到的任何语种的语言。这听起来真是太棒啦！正如经常发生的那样，昨天的科幻小说已经成为今天的科学事实。谷歌最近宣布其发明了一款名为“Pixel Buds”的耳机，这发明正是道格拉斯·亚当斯<sup>[1]</sup>（Douglas Adams）梦寐以求的。

鉴于输入了符合语法的句子，你可能会认为已经完成了驾驭语言的工作，逐字逐词对应地翻译就可以了。但是，简单的单词替换常常会把原意搅成一锅粥。以《包法利夫人》（Madame Bovary）中的一段为例：“La parole humaine est comme un chaudron fêlé où nous battons des mélodies à faire danser les ours, quand on voudrait attendrir les étoiles.”我使用法英词典逐词翻译（因为每个词有不同意思的翻译，所以必须做出取舍）：“The speech human is like a cauldron cracked where we fight of the melodies to make to dance the bears, when one would like to tenderise the stars.”（讲话人就像一口破锅，在那里，我们为了跳熊舞而战斗的旋律，当人想要软化星星的时候）。我想，这肯定不是福楼拜想要说的！这正是对于特定的语言工作机制该具有的至关重要的敏感性！当发现单词“battons”与“mélodies”连在一起出现时，我们就应该把“battons”翻译成“beat”（拍子、敲击、有规律的一连串敲打）而不是“fight”（打斗、战斗），甚至我们可以把它译成“the rhythm”（节奏、韵律）。但我仍然有一个疑惑：“tenderise the stars（软化星星）”是什么意思？

一个可以有效工作的翻译算法需要很好地理解词语之间的关联性和词语组合的模式。我记得在大学时，我最要好的哥们儿在学习波斯语。闲来无事，我随手翻看他的波斯语—英语词典时惊奇地发现，似乎每个单词都至少有三个完全不同的意思，此外单词还有阴性、阳性之分。将一个波斯语句子翻译成令人满意的英语句子是很费一番工夫的。

现代翻译算法正在深入研究语言的基本数学形态。通过试验，我们可以把语言中的单词绘制成高维几何空间中的点，然后在彼此具有结构性关系的单词之间绘制连线。例如，“men”（男人）之于“king”（国

王)的意义相当于“women”(女人)之于“queen”(王后或女王)的意义。从数学上来说,连接“men”和“king”之间的线和连接“women”和“queen”之间的线一定是平行的,并且它们指向的方向都是一致的。按照这样的方式,我们最终将得到一个高维的晶体结构,有趣的是,英语和法语的晶体结构在外观上非常近似。这是为什么?我们必须弄清楚是什么使它们保持这样相当高的近似性。

我把《包法利夫人》中的那段话放到谷歌翻译中,看看其能不能很好地捕捉到这段话的意思。谷歌翻译的结果是:“The human word is like a cracked cauldron where we beat melodies to make the bears dance, when we want to soften the stars.”(人类的语言就像一个破裂的大锅,当我们想让星星变柔和的时候,我们会敲打旋律让熊跳舞)。“soft”这个词用在这里确实要比“tenderise”好很多,但仍不够贴切。我又尝试去看OUP(牛津大学出版社)翻译出版的世界名著系列丛书中是如何翻译的,此翻译由人工完成,译者是玛格丽特·莫尔登(Margaret Mauldon)。译文是这样的:“Human speech is like a cracked kettle on which we tap crude rhythms for bears to dance to, while we long to make music that will melt the stars.”(人类的语言就像一只裂开的鼓,我们敲着粗糙的节奏让熊跟着跳舞,而我们渴望创作出能融化星星的音乐)。<sup>[2]</sup>

这个过程会让你意识到,翻译不仅要选择正确的单词,重要的是要能捕捉到字里行间的情感。算法翻译只能译出“敲打旋律让熊跳舞”,而人则能翻译出“感动天上的星辰”这样的美文。大多数情况下,机器翻译只要能传达句子的意思,大体翻译一下就可以了(诗歌不在此列<sup>[3]</sup>)。作为算法翻译成功的代表,谷歌翻译目前支持103种语言,每天翻译超过1400亿个单词。

到什么时候,从事翻译的人会失业?或是人们不再需要整篇整篇地翻译出新的文章,而是检查订正机器翻译中的小错误?对于这样的问题,我是这样看的:实际上,在人工智能解决意识的问题之前,这些算法永远不可能达到人工翻译的水平。翻译不仅仅是语言之间的迁移,而是思想在语种之间流动。除非机器有了灵魂、生出了灵智,否则它将永远无法深刻理解并充分利用人类交流的精妙。

回顾对《包法利夫人》选段的翻译,我觉得谷歌翻译的“cauldron”(大锅)比玛格丽特·莫尔登翻译的“kettle”(壶、罐、鼓)好,<sup>[4]</sup>此外,谷歌翻译的“to make the bears dance”更有冲击性和威胁

性。也许最好的翻译是人机结合的翻译。

为了获得更准确的翻译，谷歌聘请了人工翻译人员来改进它的算法，但这并不总能带来更加令人满意的结果。总是有人在试图扰乱算法，所以谷歌有时会用错误的例子加载数据，从而改变概率，最终导致翻译的错误。类似的黑客攻击也曾发生在俄罗斯，俄罗斯联邦的官方名被翻译成了乌克兰语的“Mordor（魔多）”。魔多是奇幻作家托尔金《指环王》中的地名，位于中土世界东南方，归属黑魔王索伦（Sauron）管辖。

尽管存在这些问题，谷歌翻译还是越来越擅长人类语言翻译。甚至有人提议以动物交流的声音为对象，绘制高维晶体图，看看它们的图像是否与人类交流具有相同或相似的形状，以便让我们能够理解我们的宠物到底在说些什么。当我在巴黎的索尼计算实验室见证了一个非凡的语言创造力试验（吕克·斯蒂尔斯（Luc Steels）开发了可以自行开发自己语言的机器人）以后，我开始考虑这样的一个问题：我们可能需要一种新工具来帮助我们理解机器语言了。

[1] 1952—2001，英国著名的科幻小说作家，幽默讽刺文学的代表人物，第一个成功结合喜剧和科幻的作家。此外，他也是一位广播剧作家和音乐家。其作品《银河系漫游指南》以广播剧起家，后来发展成包括五本书的“三部曲”，还被拍成了电视连续剧。——译者注

[2] 以上翻译按照英文直译，如果结合其他因素，译者认为这样更好：人类的语言只像走街串巷、撂地卖艺的马戏人在戏耍狗熊时敲打的破锣，怎能痴心妄想其去感动天上的星辰？——译者注

[3] 诗歌除了包含意思表达，还包含语音、音韵等内容，不是简单的内容翻译就可以。——译者注

[4] 译者觉得译成“破锣”更好，但这种事情见仁见智，各位读者也许有更好的翻译。——译者注

## 机器人术语

我去参观了斯蒂尔斯的实验室。镜子前20个机器人一字排开，它们探索着用自己的身体在镜子里做出各种形状。每当有新的动作和形状出现时，它们就创造一个新词来描述它。例如，机器人将其左臂置于水平位置，它就为这个动作起个名字（发明一个词）。每一台机器人都为自己独特的动作创造了一套属于自己的独特语言。

真正令人兴奋的时刻是这些机器人开始互相交流时。一个机器人从它的词典中选取一个单词，并要求另一机器人执行与该词相对应的动作。当然，第二个机器人不知道它想要什么，所以它选择一个动作作为猜测。如果猜对了，第一个机器人就会给出正向的肯定；如果没有猜对，它将向第二个机器人展示该动作。

第二个机器人可能已经给这个动作起了自己的名字，在这种情况下，它不会放弃自己的选择，而是会更新字典来包含这个新词。随着交互的进行，机器人会根据交互的成功程度来衡量单词的价值，对交互失败的单词进行降级处理，以减少交互的不畅。交互进行一周后，令人大为吃惊的是，一种共同的语言出现了。通过不断的更新和学习，机器人发展出了自己的语言。这种语言足够复杂，甚至包含了像代  
表“左”和“右”等更为抽象的单词。这些词，是在词和动作与位置的直接对应关系的基础上发展而来的。在这个试验中，任何趋同的进展都是令人兴奋的。但对我来说，真正引人注目的是：这些机器人有一种它们可以互相理解的新语言，但研究人员在试验进行的一周内一直无法理解这种语言，直到他们与机器人进行了极大量的交互，才能解译这些新词的含义。

斯蒂尔斯的试验为洛夫莱斯的预言提供了一个很好的反证。斯蒂尔斯编写的代码允许机器生成自己的语言，代码中出现了一些新的东西，演示证明除了机器以外，没有任何人类能够理解它们的共同语言。学习这门语言的唯一方法就是让机器人演示每个单词所对应的动作或位置。

谷歌的Google Brain提出了一种新的加密算法，即创建独有的语言算法，这样就可以在不被第三方窃密的情况下进行交互通信。他们在密码学的视域下进行了这样的试验：Alice的任务是向Bob发送秘密信息，Eve尝试截获并破译该秘密信息。如果Eve不能截获或破译Alice所发送

的信息，则Alice得分；反之亦然。此试验中，Alice和Bob共享一个数字，这是Eve唯一不能访问的。这个数字是Alice和Bob创建代码的关键。Alice和Bob的任务是使用这个数字创建一种秘密语言，这种秘密语言只有在知道密钥的情况下才能被解密。

一开始Alice试图掩盖信息的企图很容易被黑掉，但经过15 000次交互以后，Bob就能解读Alice所发送的信息了，而Eve解读的概率还是跟瞎猜没有任何区别。Alice和Bob使用的神经网络意味着，它们的交互很快就会被不断重新定义的语言所掩盖，所以不止Eve被挡在门外，就算是人类，即使通过查看结果代码，也不可能解读出它们正在做的交互。这些机器人可以安全地交谈，而我们人类却无法窃听它们的私语。



## 受阻于“中文房间”试验 [1]

这些驾驭着语言、把英语翻译成西班牙语、挑战并赢得《危险边缘》、理解叙事技巧的算法，提出了一个对整个人工智能领域都很重要的有趣问题：在怎样的情况下，我们应该认定算法理解它实际在做什么。我们也许可以从约翰·希尔勒设计的“中文房间”试验中得到结论。

“中文房间”的试验过程是：想象一下，你被关在一个房间里，房间里有一本用英文写成的从形式上说明中文文字句法和文法组合规则的手册，以及一大堆中文符号。依靠这本手册的说明，将中文符号组合起来，你就可以和一个讲中文普通话的人进行非常有说服力的沟通，而你不需要理解任何一个中文字或词。

希尔勒试验的目的，是为了证明这样的观点：一台被编程以文本形式进行回应的计算机，虽然我们很难将其与真正的人类区分开来，但它仍不能被认为是具有智力或理解力的。“嵌入式”这一思路是对图灵测试的巨大挑战。但是，转过来再想想，当我在说话的时候，我的大脑到底在做什么呢？这个过程中我的大脑是不是在某种程度上也在遵循一套指令？是否存在一个阈值，超过这个阈值，我们就可以认定计算机是理解中文普通话的？

当我提到“椅子”时，我知道我在说的是什么。但当一台计算机谈论一把“椅子”时，它不需要知道“椅子”这个东西是供人们坐在上面的一个物理物体。它只需要遵循规则，但遵循规则并不等同于理解。如果算法没有亲身体验过“椅子”，就不可能完美地使用“椅子”这个词。这就是为什么体现智能的问题与人工智能当前的发展趋势高度相关。

在某种程度上讲，语言就是我们周围环境的低维度投影。正如弗兰兹·卡夫卡（Franz Kafka）所言：“所有的语言都只是拙劣的翻译。”所有物理意义上的椅子都是不相同的，但它们在语言上被压缩成了一个数据点：“椅子”。这个数据点可以被另一个人打开，再将“椅子”这个数据点放到他所经历过的所有“椅子”上。有各式各样的“椅子”，它可以是扶手椅、长凳、木椅或办公椅，这些会让人产生不同的联想。这就是维特根斯坦著名的“语言游戏”。一台对具象世界没有实践经验的计算机肯定会在“中文房间”的低维空间里受阻。



归根结底，意识的奇特本质允许我们将所有信息整合到一个统一的体验中。如果我们研究一个单独的神经元，它肯定不懂英语，但当我们用一个个神经元构建出大脑中枢神经时，我们知道它确实懂得语言。当我坐在房间里用手册处理递进来的中文普通话时，我就像是大脑的一部分，是负责处理中文普通话的神经元的一个子集。你可能不明白我在说什么，换句话说，整个系统是由我、房间和手册组成的。这是整个大脑的组成，而不仅仅是我坐在那里。在“中文房间”里，我就相当于计算机的CPU（中央处理器），通过进行基本的计算来执行计算机程序的指令。

机器可以在不理解语言或不接触周围物理世界的情况下，生成有意义的句子，甚至是美的句子吗？这正是程序员们以各种方式努力解决的问题。也许机器不需要理解它在说什么，就能写出令人信服的文学作品。这又回到了让我开始对语言进行研究的问题：现代人工智能在组织语言文字来叙述故事方面能做得有多好？

[1] 也称华语房间（Chinese room）试验，是由约翰·希尔勒提出的一个思想试验，借以反驳强人工智能的观点。——译者注

## 第15章 人工智能讲故事

罗伯特·穆齐尔（Robert Musil）

追求外物真相的人会成为科学家，追求内在自我天性解放的人有可能成为作家，那二者兼具的人会成为什么样的人呢？

我从小到大读过的一些故事给我留下了很深刻的印象。罗尔德·达尔（Roald Dahl）的《达尔惊奇小说选》名列榜首，这本书中有这样的故事：一个吃了很多蜂王浆的男人变成了一只蜜蜂；一个著名的流浪艺术家拍卖自己带有文身的皮肤，价高者得；一个温良贤淑的家庭主妇，用冷冻羊腿杀死了自己的丈夫，并把羊腿交给调查此案的侦探。其中一个令人不安的故事写于1953年，讲述了“伟大的自动语法分析器”。

在这个故事中，思维呆板机械的阿道夫·奈普（Adolphe Knipe）一直想成为一名作家，可是他写出来的东西既迂腐又无趣。但后来，灵光乍现，他得到了一个启示：语言遵循语法规则，这规则的本质基本上就是数学。在这样的认识下，他开始创造一个巨大的机器——“伟大的自动语法分析器”，它能够在15分钟内模仿在世的获奖作家，写出足以以假乱真的，甚至更好的作品。奈普以他的“自动语法分析器”为要挟，威胁这些作家，让他们授权给自己使用他们的名字。在故事的结尾，讲述者与自己的良知做斗争：

此刻，坐在这里的我，听着隔壁房间里我的9个饥饿难忍的孩子在哭泣。我能感觉到我自己的手越来越靠近桌子另一边的那份“金色合同”。主啊，求你赐予我们力量，就让我们孩子继续忍饥挨饿吧。

罗尔德·达尔在这种机器被发明之前就去世了，但突然间，这似乎不再是一个疯狂的想法。

最早为计算机编写的程序是用来写情书的。图灵在布莱奇利公园破解了恩尼格玛密码机的密码<sup>[1]</sup>后，就动身前往曼彻斯特大学，将他的想法付诸实践：打造一台真正的通用型电子计算机。在他的指导下，英国皇家学会计算实验室很快生产出了世界上第一台通用型电子计算机——Ferranti Mark 1<sup>[2]</sup>。这台计算机用于寻找新的素数，解决原子理论中的问题，以及用于早期的遗传基因研究。

研究小组的成员们在实验室里发现了一些看起来像信件（或者情书）的东西，他们感到非常困惑。

### 【原文】

DUCK DUCK

you are my wistful enchantment.my passion curiously longs for your sympathetic longing.my sympathy passionately is wedded to your eager ambition.my precious charm avidly hungers for your covetous ardour.you are my eager devotion.

yours keenly

M.U.C.

### 【译文】

亲爱的宝贝儿：

你是我渴望的魔法结界。我的热情好奇地渴求着你的怜悯。我心随你意动。我珍宝一般的魅力渴求着你贪婪的热情。你是我的挚爱。

你热情的M.U.C.

M.U.C.是曼彻斯特大学计算机的缩写。克里斯托弗·斯特雷奇（Christopher Strachey）是图灵在剑桥大学国王学院的旧相识，他决定试试看Ferranti Mark 1能否更加浪漫。他根据典型的语法结构定义了通用性模板（算法）：

### 【原文】

YOU ARE MY[adjective][noun].MY[adjective][noun][adverb]  
[verbs]YOUR[adjective][noun].

### 【译文】

你是我的[形容词][名词]。我的[形容词][名词][副词][动词]你的[形容词][名词]。

斯特雷奇给计算机编写了程序，让其从自己编制的数据集中随机选取单词，并将它们插入算法的变量中。选择的随机性通过图灵创建的随机数生成器实现。在收到一两封这样的神秘情书以后，你很快会发现其中存在的一种模式，并得出这样的判断：对方一定不是一个合格的情人，因为他（她）的情书不能打动我的心。

利用算法生成文章并不是什么新鲜事。20世纪60年代，一批作家和数学家聚集在法国，使用算法生成新的作品。这群人称他们的组织为“潜在文学工作室”（Oulipo），此名称取自法语“Ouvroir de littérature potentielle”，意为“开发潜在的文学”。其创始人之一的雷蒙·格诺

（Raymond Queneau）认为：“规则的约束是创意过程的重要组成部分。盲目地跟从每一种冲动所带来的灵感，实际上才是一种真正的奴役。通过对文学创作施加准数学的规则约束，你将获得一种新的自由。”该组织早期的项目对象主要是诗歌。写过诗的人都知道，相对于自由形式的散文，诗歌的规则约束常常会迫使你用新的方式、方法来表达思想。

这个组织最受欢迎的算法，是由让·莱斯库尔（Jean Lescure）构思出来的“S+7模式”（也就是英语的“N+7模式”）。“S”取自法语的“substantifs”，意为“实体的”，即名词。该算法将任意一首诗作为输入，然后对诗中的所有名词进行操作，将其替换为该词在字典中所在位置向前或后移动7位的单词。输出的则是原诗的改写版本。下面以布莱克（Blake）的诗为例。<sup>[3]</sup>

### 【原文】

To see a World in a Grain of Sand

And a Heaven in a Wild Flower

Hold Infinity in the palm of your hand

And Eternity in an hour

### 【译文】

一沙一世界

一花一天堂

无限掌中置

刹那成永恒

改写后

### 【原文】

To see a Worm in a Grampus of Sandblast

And a Hebe in a Wild Flu

Hold Inflow in the palsy of your hangar

And Ethos in an housefly

### 【译文】

喷砂现蠕虫

流感见青春

麻痹入机库

家蝇精神足

莱斯库尔希望通过这个有趣的试验，能促使我们用新的眼光来重新审视原文。算法替换了名词，但保留了句子的基本结构，它可能有助于显示出被词意所掩盖的语言结构。

作为法国数学学会的一员，格诺不仅研究数学，还研究过哲学，他对数学和创造力之间的联系非常感兴趣。他以数学为工具，尝试利用很多不同的方法来创作新诗。在“潜在文学工作室”成立前不久，他就写成了一本十四行诗集——《100 000 000 000 000首诗》。他写的诗每行有10种不同的版本：如果第一行有10种版本，第二行也有10种版本，那么前两行就可以构造100种版本。依此类推，一首十四行诗就会有10<sup>14</sup>种版本，也就是整整100万亿首诗！如果侏罗纪时期进化出来的第一只梁龙，以每分钟一首的速度背诵，那么直到现在，它才刚刚背完一遍。

格诺创作出了文学版的“莫扎特骰子游戏”，有可能是下面这首十四行诗<sup>[4]</sup>。这是我随机挑选的，它从未出版过：

### 【原文】

Don Pedro from his shirt has washed the fleas  
His nasal ecstasy beats best Cologne  
His toga rumpled high above his knees  
While sharks to let's say potted shrimps are prone  
Old Galileo's Pisan offerings  
Nought can the mouse's timid nibbling stave  
He's gone to London how the echo rings  
The nicest kids for stickiest toffees crave  
Emboggled minds may puff and blow and guess  
In Indian summers Englishmen drink grog  
And played their mountain croquet jungle chess  
We'll suffocate before the epilogue  
Poor reader smile before your lips go numb  
Fried grilled black pudding's still the world's best yum

正如“潜在文学工作室”的实践所表明的，诗歌是一种特别适合应用算法的文体。模式规则的约束性确定了一个模板，算法可以尝试以一种有意义的方式填充该模板。对于俳句或十四行诗，它们只是不同的模式，算法的任务是选择与模式匹配的单词，同时尝试寻求某种形式以便使其整体连贯。每当写诗需要押韵时，我发现押韵词数据库是超级有用的。通过控制押韵和限制节奏来编织一行诗，这对于计算机来说完全不



是什么事儿。

这也就是“控制论诗人”（Cybernetic Poet）算法代码背后的基本原理。“控制论诗人”是未来主义作家雷·库兹韦尔（Ray Kurzweil）的最新作品，他经常撰写关于即将到来的“人与机器相互融合”的文章。库兹韦尔从不依赖字典中随机挑选出来的词，而是训练他的“控制论诗人”学习像雪莱和艾略特这样有成就的诗人的作品。下面是“控制论诗人”的一首三行俳句诗，由济慈的一段经文改编而成：

【原文】

You broke my soul

The juice of eternity

The spirit of my lips

【译文】

你突破了我的灵魂

永恒的汁液

我唇齿间的精灵 <sup>[5]</sup>

虽然这首诗确实是有17个音节，但是算法似乎忽略了俳句诗应该遵循的对称性原则：第一行5音节，第二行7音节，第三行5音节。

下面引用一首融合了雪莱和艾略特风格的诗：

【原文】

Lady of Autumn's being

Thou, from the day, having to care

Teach us now thoroughly small and create

And then presume?

And this, and me

And place of the unspoken word, the unread vision in Baiae's bay

And the posterity of Michelangelo

### 【译文】

秋之女神

汝从天而降，关怀万物

让我们知道自己的渺小，教会我们如何创造

这一切理所当然吗？

涵盖一切，包括我

从贝亚湾未见未闻之处

到米开朗基罗的子孙

《西风颂》与《普鲁弗洛克的情歌》在这首诗里相遇。

库兹韦尔所进行的图灵测试中，“控制论诗人”能够在大多数情况下骗过人类评委。这在一定程度上是因为，精辟的作品是现代诗歌的一部分，现代诗歌留下大量解读的工作给读者去做。所以，算法生成高深莫测的诗作足以以假乱真。如果你对这些诗歌有兴趣，可以前往库兹韦尔的网站 [6] 进行浏览。

如果你对辨别诗歌到底是人类写作的还是机器算法生成的很感兴趣，可以看看本杰明·莱尔德（Benjamin Laird）和奥斯卡·施瓦茨（Oscar Schwartz）的网站 [7]。在这个网站上他们推出了一个颇具挑战性的诗歌图灵测试项目——“是人，还是机器人？”（bot or not）。

“控制论诗人”可能在创作足以以假乱真的诗歌方面做得很好，但要创造一个“控制论小说家”所要面对的挑战则难得多。

[1] 1939年9月4日，即英国对德国宣战的第二天，图灵前往布莱奇利公

园（即战时政府信号密码学校，政府通信总部的前身）报到，为英国进行密码破译和教学等工作。恩尼格玛密码机（德语：Enigma，又译为“哑谜机”，或“谜式密码机”）是一种用于加密与解密文件的密码机。确切地说，恩尼格玛密码机是对二战时期纳粹德国使用的一系列相似的转子机械加解密机器的统称，它包括许多不同的型号，为密码学对称加密算法的流加密。——译者注

[2] 也有人称世界上第一台通用型电子计算机是埃尼阿克（ENIAC）。1946年2月14日，它诞生于美国宾夕法尼亚大学，研究人员于次日正式对外公布研发成功。——译者注

[3] 玩味英文更有意思。译者才疏学浅，在这里只能将原文强译成中文，望各位读者谅解。——译者注

[4] 十四行诗是欧洲一种格律严谨的抒情诗体，翻译成中文会破坏格律、音韵等，所以此处不翻译，请各位读者自行体会。——译者注

[5] 这里只是意译，仅供参考。后文讨论英文语音等问题时，请参照原文。——译者注

[6] 网址：<http://www.kurzweilcyberart.com/>。

[7] 网址：<http://botpoet.com>。

## 如何在一个月内完成一部小说

作为对“全国小说写作月”<sup>[1]</sup>（National Novel Writing Month）的回应，莱斯库尔想将算法作为写作现代文的一个工具来使用，所以他组织了“全国小说生成月”（National Novel Generation Month, NaNoGenMo）活动，并成功吸引了众多的程序员参加。软件开发员兼艺术家达赖厄斯·卡泽米（Darius Kazemi）觉得，与其每天绞尽脑汁才只能写出1667个单词，还不如花一个月的时间写代码，然后瞬间生成一部5万字的小小说。他还计划把生成的小小说和生成小说的代码都分享给大家。2013年，他在Twitter上发布了他的想法，同时发起了每年一次的“年度编程文学马拉松”（Annual Literary Hackathon）大赛。

参与NaNoGenMo活动的程序员大多数采用“扰动”现有文学作品的思路生成小说。例如，借助“Twitter过滤器”改写《傲慢与偏见》；利用“科幻算法”重新演绎小说《白鲸》；通过代码重新诠释古斯塔夫斯·辛德曼·米勒（Gustavus Hindman Miller）的经典作品《一万个梦》。一部构思非常大的作品《探索者》（seeker）引起了人们的广泛关注。生成此小说的算法通过阅读wikiHow<sup>[2]</sup>上的文章，努力理解人类的大脑如何“运转”。该算法的元代码被划分为操作、浏览、联想、再现四个环节。程序“thricedotted”的开发者向我们解释了各环节需要完成的工作：

在操作环节，程序阅读某文章并将其中涉及人类活动的概念剔除。其以上一环节得到的“种子概念”为基础，搜索纯文本的“记忆”，然后以浏览模式中无法识别的概念（已删除的记录），围绕“种子概念”进行“不可见”的联想。

《探索者》在探索wikiHow数据库的过程中，也记录了算法的发现之旅：从无知到某种表象程度上的理解。它浏览的第一个“如何……”的页面是“如何让女孩提出与你约会”。其从中获得一个种子概念——“伤害”，涉及怎样做才能不伤害到一个女孩的感情。之后算法围绕着“伤害”展开联想，即兴生成了超现实主义的文字段落。

与其他由算法生成的小小说不同，《探索者》几乎可以说是成功的：你会感觉自己开始进入机器的“大脑”，因为它试图去理解人类。虽说该小说读起来像是一串奇怪的计算机代码，但这与算法内在的思想相一致。这或许就是机器生成文学作品的终极目标：理解一种全新的意识形

态，并发现其与人类自我意识的不同之处。

目前，对于算法的商业用途而言，可以战胜《拯救灰姑娘》（Mills&Boon Romance）或丹·布朗（Dan Brown）的冒险小说就已经足够好了。在这些畅销书中，许多都基于明确的规则或套路。这个过程难道不能自动实现吗？如果算法不能生成伟大的文学作品，或许它们可以生成肯·福莱特所写的商业图书，甚至是算法版的《五十度灰》（Fifty Shades of Grey）。组稿编辑朱迪·阿彻（Jodie Archer）和数据分析师马修·乔克斯（Matthew Jockers）合作编写的算法，可以判断出一本书能否成为畅销书。该算法发现，就读者的习惯而言，相较于文学小说的读者，畅销书的读者更喜欢：短小精悍的句子，强而有力的具有推动性的语言，以对话形式进行的叙事方式以及浅显易懂的用词。要是我在开始写这本书之前能知道这一点就好了！

[1] 每年11月，作品字数超过50 000字即可参加。——译者注

[2] 该网站旨在建立全世界最大的最高质量的指导手册数据库。无论你想做什么，该网站拥有的多语种指导手册都可以为你提供免费的逐步指导。——译者注

## 哈利·波特和死亡波特尼克 [1]

到目前为止，我提到的大多数例子都基于自上而下的编程思想。例如，按照明确的规则，可随机填充单词的诗歌模板；将经典作品转换为新作品的代码；通过编程获取数据并将其转化为故事的算法。这些程序实际上并没有太多自由发挥的空间，但机器学习正在改变这一切：算法可以从小说作者的所有作品中学习，探索他们的写作方式。如果他们喜欢使用某个特定的词，这个词后面很可能会跟着其他出现频次高的词。通过构建作者使用单词的概率图，算法可以生成该作者作品的续篇。这就是预测性文本生成器的工作原理，其产出的文学作品既发人深省又引人入胜。

一个名为“Botnik”的公司倡导使用机器学习来创造新文学。该公司由作家杰米·布鲁（Jamie Brew）和《纽约客》（The New Yorker）前漫画编辑鲍勃·曼考夫（Bob Mankoff）于2016年创立。现在其已发展成为一个开放性社区，倡导作家借助科技手段创作喜剧。该公司有个同名产品，就叫作“Botnik”，它是一款新的人工智能幽默应用程序。他们将美国喜剧《宋飞正传》（Seinfeld）的剧本作为Botnik的学习对象，对其中的对话进行了数学分析并制作了新的剧集。他们甚至还请到了《实习医生风云》中的演员扎克·布拉夫（Zach Braff）出演了一集由Botnik根据该医疗喜剧改编的独角戏。算法有时会输出出人意料的超现实主义作品，比如Botnik为《宋飞正传》写出了精彩的剧本：“约会的对立面是金枪鱼，而三文鱼的对立面是任何其他事物。我相信你明白我的意思。”Botnik还制作了一份感恩节食谱，并在YouTube上发布了一段视频，让人们了解如果把烹饪这件事交给算法，你将会吃到什么样的感恩节特别晚餐：准备感恩节食物的最好的方法就是把火鸡“对折”，然后直接扔进厨房。

截至目前，他们最成功的作品可能要算Botnik在学习了《哈利·波特》七卷丛书之后，生成了三页几乎可以以假乱真的续集。可以说，有些时候，算法才是最纯粹的“天才”。

### 【原文】

Magic: it was something that Harry Potter thought was very good. Leathery sheets of rain lashed at Harry's ghost as he walked across the



grounds towards the castle.

Ron was standing there and doing a kind of frenzied tap dance. He saw Harry and immediately began to eat Hermione's family. Ron's Ron shirt was just as bad as Ron himself.

### 【译文】

魔法——哈利一直认为这是一种很好的东西。哈利穿过地面朝城堡走去时，皮料一般密布的雨帘猛烈地鞭打着他的鬼魂。

罗恩站在那里，发疯似的跳着踢踏舞。他看到哈利，立刻开始吃赫敏的家人。罗恩的罗恩衫就像他自己一样糟糕。

由于故事情节的发展过分自由，导致了续写剧情不太可能超过三页。但我想，对于那些迫切希望从魔法世界中了解更多的忠实粉丝来说，这至少聊胜于无。

我决定做一个试验：如果把我的第一本书——《素数的音乐》（The Music of the Primes）作为Botnik的输入，它能否给我一个独特的新视角。但我却得到了以下奇怪的输出：

### 【原文】

The primes are the jewels which shine amongst the vast expanse of our infinite universe of numbers. As he counted higher and higher Gauss suddenly saw a pattern beginning to emerge. His passion for the problem was further fuelled when his father offered to buy him a Ferrari. Previously education schemes had been geared to the creation of each list of primes 2, 3, 5, 7, 11 and 13 years respectively. For all but their last year they remain in the ground feeding on the sap of tree roots.

### 【译文】

素数是一颗在我们广袤无垠的数字宇宙中发出耀眼光芒的宝石。数着数着，突然间高斯发现了一个规律。在他的爸爸答应给他买一辆法拉利时，高斯探索这个问题的热情更加高涨了。以前的教育计划都是按照清单针对每个素数分别为2年、3年、5年、7年、11年和13年。除了最后

一年，它们都留在地里，以树根的汁液为食。

根据这个输出的结果，可以辨认出这就是我的书，但它呈现出一种奇怪的表达。从这次试验中我得出一个重要经验，那就是在生成文本时仍然需要大量的人工干预。算法所做的是为你提供一个18个单词的选择范围，这些选项基于目前已存在的数据，这些单词会尽可能地遵循这些数据。这样就给了我很大的自由，让我可以随心所欲地把它带到我的创作想象当中。通过算法进行艺术创作时，人的影响成分通常会被掩盖。比如，算法生成了一个更好的作品，大家会说“人工智能写出了新的《哈利·波特》”，而不是“又一个文学生写了一本新小说”。

坦白地讲，我觉得小说家们很快就会失业是不大可能的。也许Botnik确实可以从构建句子的方式中学习、辨认、识记，但其所捕捉到的只是这样一个事实存在：作者的写作风格。它只捕捉到了这一点——文本的局部，却没有从整体上考虑叙事结构。就像爵士乐的“续作者”一样，它可以生成一些以假乱真的乐句，但最终还是难逃走向乏味、无聊和枯竭，因为算法不知道该继续向何处发展音乐。我常常怀疑奈飞、亚马逊等公司是不是已经在使用这些算法，用来淘汰那些我们一直在看，却从中一无所获的剧本。

[1] 波特尼克为Botnik的音译，这是一家研发人工智能自然语言技术的公司，2016年在美国成立。其致力于开发能与人协作的内容创作工具，即用人工智能挖掘人类的创造潜能。此标题应为致敬《哈利·波特与死亡圣器》。——译者注

## 假如.....

2012年，乔治亚理工学院的马克·里德尔（Mark Riedl）和他的同事开发了一个会讲故事的算法“Scheherazade-IF”，旨在解决上述问题：它的目标是在故事发展可能性的迷宫中找到一条更加连贯畅通的路径。这个算法名字中的“Scheherazade”来源于王后山鲁佐德为暴虐的国王讲故事来安抚他的传说，她是一位优秀的故事讲述者。Scheherazade-IF中的“IF”取自“Interactive Fiction”，意为交互式小说。如果你要求Scheherazade-IF构建一个关于某个特定主题或情境的故事，而这个主题或情境它之前从未遇到过，那么它就会通过搜集和消化之前的故事来进行学习。

该算法的主要开发者之一里德尔表示，由于人类具备许多现实世界的知识，所以他们都是天生的、优秀的故事讲述者。Scheherazade-IF从由一大群人构成的大规模知识库中汲取新的信息并充分消化，再将这些范例汇总形成一个树状图，图中列出故事发展的各种可能方向。好的故事讲述者，会在故事发展树状图中找到最佳的路径。在很多故事情节会随玩家的选择而走向不同结局的开放性电脑游戏中，这种方法非常奏效。

这里介绍一种我孩童时期非常喜欢的故事叙述方式。在读《游戏书》（Gamebooks）或《选择你自己的冒险旅程》（Choose Your Own Adventure）时，你会遇到一些需要做选择的环节：如果你选择左侧的门，请跳到第35页；如果你选择右侧的门，请跳到第39页。有时你的选择会让故事的发展产生比较严重的不连贯和不合逻辑的结果，因为即便是一个只有10个结点的故事，也可以产生超过1000种不同的情节发展方式。这就需要有一种方法，可以让算法从中选择出最精彩的那一个。

Scheherazade-IF正尝试通过它在网上收集到的数据，生成故事发展的可能性树状图，进而实现这一点。那么，它选择的“令人满意的路径”，到底能获得多高的满意度呢？研究团队的测试表明，它所做的选择获得的满意度得分可以与人类相媲美，远远高于随机生成的故事。同时，由该算法产生的逻辑性错误的次数明显少得多。比如，在第2章中就死去的角色，在第5章又突然出现了。除非这是一个僵尸故事，否则类似的这种错误一旦出现，便会立刻暴露无遗、无所遁形。

从网上搜罗以前的故事把它们重新组合在一起固然不错，但想象一下如果是前所未有的场景会有怎样的结果？这也是由欧盟资助的Whim（What If Machine）项目所试图解决的问题。作家在尝试创造新作品时面临的一个问题是，他们会卡在自己有限的固定思维模式里。Whim通过提供焕然一新的故事场景，把故事讲述者带出他们的舒适区。

每当我们想要构思一个新故事时，经常会做一些假设：假如一匹马能够飞翔，就诞生了“飞马”（pegasus）；假如一幅画像中的男子会逐渐变得衰老和丑陋，而他本人却一直保持着年轻俊美，就有了道林·格雷（Dorian Gray）这一角色的塑造；假如一个女孩贸然闯入一个陌生的地方，那里的动物都会说话，那里的人都有着奇怪的行为举止，就有了《爱丽丝梦游仙境》的奇幻之旅。我小时候非常喜欢的《达尔惊奇小说选》中的许多故事都是在“假如……”的创造力模型基础上构想出来的。

事实上，人类讲故事可能起源于“假如……”这种问答结构。讲故事是人类做安全试验的一种方式。如果你问一个人“假如……”，其实是在探索你的行为对他可能带来的影响。最早出现的故事极有可能就源自我们对在周遭混乱的环境中寻找某种秩序的渴望，想要在这个残酷无情的宇宙中找到自身存在的意义。这是一种早期的科学形式。围坐在火堆旁分享当天的狩猎故事，可以帮助部落在第二天能获取更多的猎物。智人力量上的短板，在部落的集体智慧中得到了弥补。这种智慧的力量随着社交和分享的增加而增强。营地的篝火点燃了人类创造力的火种。

Whim项目旨在引燃数码科技的创造力篝火，它最初就是以飞马为出发点的：一匹会飞的马。算法能不能构想出其他有趣的动物以便推动故事情节的发展？先从一个动物数据库开始，该数据库列出了动物们拥有的所有习性。国家地理儿童版网站就是一个很好的起点，它告诉我们海豚是一种生活在海里的哺乳动物，人类甚至可以骑在它的身上；鸚鵡是一种会飞会唱的鸟。通过对这些特性的组合和匹配，可能会孕育出一种既会飞又会歌唱的哺乳动物，你可以骑着它。这种新生物出现在童话或《哈利·波特》丛书中一点都不足为奇。

就像那些长有头、躯干和腿的动物一样，将不同事物的组成部分解构，然后再通过混合进行重构，就可以创造出千奇百怪的新事物。假设动物的每个身体部位都有10种选择，那么10个身体部位就会创造出1000种不同的生物。如果组合有效，那么就需要有评估这些组合的办法。Whim团队引入了对创意的新颖性和刺激程度进行评价的数学函数，并

以此过滤那些含糊不清、毫无意义的构想。最后他们得到了这样一些有趣的组合：凭借眼睛来保护自己的生物，长着翅膀的飞虎，生活在森林中但又能在水里游泳的鸟。

这些具有奇技异能的新生物是促使新故事诞生的极佳催化剂。下一步就是通过编程使Whim生成新颖的叙事思维：从一系列的“假如.....”开始，随即将故事情节展开，再打乱这些场景中隐含的各种假设，最后以一种令人惊讶的、颠覆性的方式与主题相结合。通过这样的方式可以激发创造力。Whim被设定在六个虚构的场景类别基础上生成叙事建议，分别是：卡夫卡、抉择、乌托邦和反乌托邦、隐喻、音乐剧和迪士尼。它们都成功生成了各不相同的输出结果。

在迪士尼场景中，Whim构思出了这样一个故事情节：假如一个小原子失去了它的中性电荷.....这一桥段或许会出现在下一部的《头脑特工队》当中，又或许就是我们身边的科学极客们头脑中的一个新的想法。该场景生成的另外一些建议可能更应该被归类在反乌托邦场景中，比如：“假如一架小飞机找不到机场.....”

抉择场景生成的故事情节成功的可能显然不太高：“假如一台旧冰箱找不到足够坚固的房子.....与之相反，它发现了一种以水做成的特殊风格的雕像，于是就再也不想要坚固的房子了。”卡夫卡场景亦是如此：“假如有一辆自行车出现在狗栏里，突然间这辆自行车就变成了一只只会开汽车的狗.....”

由Whim生成的一个故事情节，最终促成了一部音乐剧于2016年在伦敦西区上演。天空艺术电视频道（TV channel Sky Arts）对探索算法创造力的极限非常感兴趣，所以其委托人工智能制作了一部音乐剧。电视台对该音乐剧的全部制作过程进行了跟踪录制，直至其被搬上舞台。为了给这部音乐剧构思一个场景，Whim团队也加入进来。算法提出了一系列不同的场景，然后由剑桥大学开发的另一种算法进行筛选。筛选算法分析音乐剧走红或惨败的原因，为Whim提供故事情节进一步发展的建议。比如，很可能会成为热点的故事情节有：假如一名伤痕累累的士兵想要获得真爱，就必须先得学习去理解和体会一个孩子的内心.....

在生成童话故事方面，另一个算法“普洛普写手”（PropperWryter）更胜一筹。1928年，形式主义语言学家、民俗学家弗拉基米尔·普洛普（Vladimir Propp）在他的《故事形态学》里提出，俄国民间故事有31种叙事原型。“普洛普写手”把Whim提供的情节最终发展成为一个关于

格林汉姆普通妇女反核运动的故事。其音乐由另一个名为“机器人劳埃德·韦伯”（Android Lloyd Webber）的算法提供。

2016年春天，《飞越藩篱》在伦敦西区的艺术剧院进行了短暂的演出。为完成这一作品，人类的干预可能和电脑的创造力一样多。该剧的上演对于安德鲁·劳埃德·韦伯并没有造成太大的影响。戏剧评论家琳恩·加德纳（Lyn Gardner）对这部音乐剧给出了2星的评价，评语如下：“这是一部过时的、中庸的音乐剧，剧中充满了令人愉悦的中庸的歌曲，还有可笑的老套场景和角色。”但也许真正值得我们反思的是，审稿人并没有给予算法太多的信任。



## 伟大的“自动化”数学家

数学家如果能提出诸如“假如.....”这一类的问题，就意味着他距离突破知识边界已经只有一步之遥了。比如：假如有一个数的平方值为-1；假如弯曲空间中，两条平行线可以相交<sup>[1]</sup>。打破固有的结构框架，从变化中探索有价值的新发现，是编写新的数学“故事”的经典工具。数学中的“假如.....”真的有助于创造新的数学吗？如果数学是用数字讲述的故事，那么目前的算法在生成新的数学故事方面有多大的价值呢？

西蒙·科尔顿是“绘画傻瓜”代码的编写者，也是Whim项目的协调人。他与伦敦帝国理工学院（Imperial College London）的斯蒂芬·马格尔顿（Stephen Muggleton）联手对上述问题进行了探索。他们开发的算法基于现在已经公认的数学成果，验证能否激发出新的想法。科尔顿在访问量最大的数学网站之一“整数序列在线百科全书”<sup>[2]</sup>（The On-line Encyclopedia of Integer Sequences）上发布了这一算法。该网站的发起人是尼尔·斯洛恩（Neil Sloane），其目的旨在收集所有有趣的数列，并探索生成这些数列的公式或方法。该网站包括一些经典的数列，比如：

1, 1, 2, 3, 5, 8, 13, 21.....

所有读过《达·芬奇密码》的人都会认出这就是著名的斐波纳契数列——数列中的每个数都是由其前面的两个数字相加而生成的。又如：

1, 3, 6, 10, 15, 21.....

这个数列被称为三角数，即正整数前n项之和构成的数列。

你还会发现数学书籍中最神秘的序列之一：

2, 3, 5, 7, 11, 13.....

这是由素数（或者说除了1和它本身以外不再有其他因数的自然数）构成的数列。对于该数列，很难找到一个通项公式来生成下一个数，这也是数学界公认的未解之谜之一。如果哪天机器算法攻破了这一难题，我想我们就都可以卷铺盖回家了。

整数序列在线百科全书数据库也包含了令我如醉如痴的数列，即编

号为A158079的数列：

1, 2, 5, 15, 67, 504, 9310.....

这些数字分别是31、32、33、34、35、36、37阶对称群拥有的对称性元素个数。我的研究表明，它们遵循类似斐波纳契数列的规则，但我仍在这些已有的数字中寻找一种特定的组合方式，以便得到该数列的下一个数字。

科尔顿决定用他的算法来实现对新数列的识别和解释。科尔顿的同事托比·沃尔什（Toby Walsh）提出了一个名为“可重分解因子数”<sup>[3]</sup>

（refactorable numbers）的概念，其定义为：数列中每个数字的因数个数，本身又是该数字的因数（比如，数字9的因数有3个，而3本身又是9的因数之一）。该数列中的奇数，被称为“奇数可重分解因子数”<sup>[4]</sup>

（odd refactorable numbers，这个叫法听起来或许有些奇怪）。算法推测：所有的奇数可重分解因子数都是完全平方数。虽然没能证明这一点，但算法提供的这个建议已然足够引起科尔顿的兴趣。他证明了这命题，并发表了一篇期刊论文来解释具体证明的过程。可以说，可重分解因子数是一项由机器生成的发明。它是伟大的“自动化”数学家这颗新星从地平线上冉冉升起的第一个迹象吗？

[1] 黎曼几何研究的是一个弯曲的空间，其中的直线并不是我们通常所说的直线。比如在球面几何上，两条经线是平行的，但直观上它们是相交的。——译者注

[2] <http://oeis.org/>。——译者注

[3] 在整数序列在线百科全书数据库中的编号为A033950。——译者注

[4] 在整数序列在线百科全书数据库中的编号为A036896。——译者注

## 人工智能新闻

人工智能新闻报道算法的核心是如何将未经处理的原始数据转换成新闻报道。遍布世界各地的各大公司每周都会发布有关收益的数据。过去，像美国联合通讯社（以下简称“美联社”）这样的新闻机构必须指派一大批记者去仔细研究这些数据，然后编写一份关于这些公司经营状况的报告。这项工作非常烦琐和乏味，效率很低。一年之内，你只能覆盖约1000家公司，这意味着很多人们感兴趣的公司没有被报道。办公室里的记者都很害怕被选中写这些报道，它们成了记者的梦魇。

因此，很少有记者为美联社决定使用机器来帮助报道这些新闻持反对意见。像“Wordsmith”这样的算法，具有自动化的洞察力、科学的叙事能力，现在正被应用于基于大量数据的分析报道的撰写工作。相较于过去的人工撰写，它真是再合适不过了。在大部分情况下，只有读到文章末尾处的署名时，你才会发现这篇文章是机器写的。这些算法解放了记者，让他们可以去撰写更重要的新闻。

数据挖掘算法对美联社的报道很重要，对企业来说也越来越有用。算法可以获取大量的商业信息，并将不能直观读取的电子表格转换成即使是普通的公司职员也能够理解的新闻报道。算法不仅可以统计出公司或员工的工作绩效数据，还可以依据月度数据的细微变化进行预测。这些隐藏在电子表格或各种图表中的细微差别很容易让人们忽略，但一旦它们被机器以自然语言的形式解释出来，就会很容易引起共鸣。对于希望把握公司潜在估值和变化的投资者来说，这些信息描述尤为重要。

此外，这些算法还会参考读者的阅读喜好，生成那些像是蹲在家里就能写出来的、深受大众喜爱的、在小报封底上读到的那种武断又充满讽刺意味的体育新闻，或者带有政治偏见的报道。以地方新闻机构拥有的记者人数来看，其覆盖不了所有的比赛，因此越来越多的记者开始使用算法将足球或棒球比赛的结果变成可读性强的新闻报道。除此之外，曾经对机器替代自己工作而感到恐慌和震惊的一些记者也开始尝试借助算法去生成清晰的新闻报道。有这样一个例子，乔治·华盛顿大学体育网站上的一篇报道中，并没有赞颂对方投手的表现，哪怕是他拥有9局淘汰了27名击球手，确保无人上垒的完美战绩。这让记者们认为，由于这一结果太过罕见，算法无法通过编程来进行报道。

但事实上，这篇文章是由人撰写的。他本身可能支持惨败的球队，所以就带有倾向性地将比赛结果隐藏在文章的倒数第二段。研究叙事科学性的团队饶有兴致地将此次比赛结果作为参数输入算法，想要看看它会如何处理。以下内容是仅依靠文中给出的数据生成的文章开头部分：

星期二对于罗伯茨来说是一个伟大的日子，这位年轻的投手在比赛中表现完美，帮助弗吉尼亚队在达文波特球场以2比0战胜乔治·华盛顿队。27个主场队员来到棒球场上，这个弗吉尼亚投手战胜了他们，在比赛中表现完美。在这场由他掌控的令人难忘的比赛中，他击败了10个击球手，让瑞恩·托马斯在决赛中出局。

此次报道，算法得1分，记者得0分。

除了现实生活中的体育赛事，越来越多的新世代球迷开始热衷于在梦幻棒球游戏（Fantasy Baseball）中组建自己的虚拟球队。在美国和加拿大，有近6000万人选择游戏中的美国职业橄榄球大联盟（NFL）的虚拟球员组成球队，与他们的朋友比赛。他们平均每年要花29个小时来管理自己的球队。雅虎已经开始使用“Wordsmith”从每周生成的NFL虚拟球队数据中生成个性化的新闻。人类不可能花费大量的时间，每周创造出数百万条新闻，来满足游戏玩家了解自己的虚拟球队在比赛中的表现的欲望。

当然，任何事物都有两面性。通过算法生成的新闻报道获取信息也有不好的一面。历史一再提醒我们，故事是一种强大的政治工具。近期的研究告诉我们，数据和证据很少能改变人们的想法，只有当它们被编织成一个故事时，才有足够的说服力去改变人们的想法。就像某些觉得给孩子接种疫苗会非常危险的人，很少会被疫苗能够有效预防疾病传播的统计数据说服。但是当你给他们讲了一个人因为没有注射疫苗而得了麻疹或天花的故事，再把这个故事和数据结合起来，就有可能引起他们的重视。正如乔治·蒙贝尔特（George Monbiot）在《走出废墟》（Out of the Wreckage）中所说的：“取代一个故事唯一的办法，就是讲述另一个故事。”

像剑桥分析公司（Cambridge Analytica）这样的企业也已经开始“无情地”利用新闻故事来改变人们的观点了。他们通过一款名为“这就是你的数字生活”（This is your digital life）的应用程序，收集了8700万位Facebook用户的个人信息，用于开发人的“心理学档案”，然后通过新闻报道进行比对，影响选民在选举中的抉择。起初该算法随机分配新闻

故事，但它会逐渐了解到哪些故事可以吸引更多的点击量。

他们很快发现，美国那些思想保守的白种年轻人对“抽干沼泽”“筑起高墙”之类的短语反应积极。因此，算法将自己生成的这类故事推送到他们的Facebook页面，以满足其对这类新闻的胃口。算法所做的，就是确保这些故事能出现在最可能受其影响而改变观点的人面前，而不是浪费在其他人的身上。

剑桥分析公司实际上操纵了选民。该消息传出后，公众的强烈反应令该公司陷入困境。更具讽刺意味的是，这恰恰揭示了该公司的宗旨和核心价值——新闻故事对事件的影响力。

尽管剑桥分析公司已经倒闭，但仍有许多公司在继续挖掘数据，为那些愿意付钱的机构或个人提供战略决策。值得注意的是，倘若我们想要为生活保留一点控制权，就需要了解我们的情绪和政治观点是如何被这些算法摆布的，以便在外部信息相同的情况下，能够根据自己的情况去进行判断。

在这里我必须承认一点，这本书不全是我自己写的。我向罗尔德·达尔的“伟大的自动语法分析器”的最新版本“屈服了”。这本书有一个350个单词的段落是由算法编写的，该算法可根据输入的关键字自动生成短篇文字。它通过文学版图灵测试了吗？你有注意到吗？<sup>[1]</sup>

允许算法基于现有文本生成文章所带来的风险，当然就是剽窃。这个算法可能会给我带来麻烦。我设法在网上追溯，终于在另一个网站上找到了一篇文章，与它生成给我的段落有一些惊人的相似之处。倘若因为剽窃文章而被原作者起诉，人们就会觉得人工智能生成的文章并不像想象的那么好了。

尽管算法具有可变性和创新性，但目前就讲故事的算法而言，它们并不会威胁到作者——“伟大的自动语法分析器”仍然是人类的幻想。即便是数学家之间讲的逻辑故事，也是人类思维的专属品。既然有这么多故事可以讲，那么选择哪些故事仍然是一个问题。只有人类创造者才会明白为什么另一个人的大脑会跟随他们的创造之旅。毫无疑问，计算机一定在创造的旅途中为我们提供帮助，但它可以成为的是望远镜或者打字机，而不是故事的讲述者。

[1] 对于读者来说，本书中文版已经过了翻译，故无迹可循了。——译

者注



## 第16章 为什么要创造：思想的交流

道格拉斯·霍夫斯塔特

创造的精髓无关机械，但每一个创造的结果都需要机械来实现，这解释起来并不会比解释打嗝更简单。

计算机是扩展人类智慧的强大工具。在围棋比赛中，它帮助我们发现了新下法，扩展了棋路。爵士音乐家们从算法生成的爵士乐中听到了在他们潜意识中出现但并未实现的音乐。以往人类无法驾驭的数学定理，现在依靠算法也变得唾手可得。生成式对抗网络算法已能创造出可与国际艺术博览会展出的艺术作品相媲美的作品。回顾我的研究生涯，至少目前计算机还没有产生对人类创造力构成威胁的任何东西。

在我的经历中，我一直都举棋不定，我曾经坚信算法永远不会生成水平接近人类所创作的绘画、音乐、文学作品，也曾意识到艺术家的行为在某种程度上是依赖于其自身的“算法”对周围世界所做出的响应。机器能否轻易拥有依赖人类思维和智慧所产生的丰富而复杂的响应能力？人类的思维和智慧已经进化了数百万年，问题是，这种进化能以多快的加速度发展？

我认为机器学习的新思想挑战了许多关于机器永远不可能具有创造性的传统论点。机器学习不需要程序员理解巴赫是怎样创作众赞歌的，因为算法可以自行获取数据并学习。这样的学习使我们对人类艺术家的创作过程有了新的认识。但这种“创造”是否只会带来更多的“重复物”？算法是如何从正在学习的数据中跳出来看问题的？即便有许多疑问，我们还是能看到在艺术世界中还有开发出未被触及的新领域的可能性。例如，爵士音乐家接纳算法生成的曲子作为他音乐的一部分，最终为他带来了一种即兴演奏的新方法。

许多人会认同，探索型创新和组合型创新可能是算法能够实现的，因为其依赖于人类早前的创造力，算法会扩展或组合这些创造力。但传统观念认为，用算法产生变革型创造力似乎是不可能的：一个被束缚于系统内部的算法怎能找到突破自我的方法，进而做出一些令人惊喜的事情？人工智能新的应用展现了我们如何创建打破规则束缚的元算法以及后续会产生什么效果——变革型创造力并不是无中生有，实际上它是

在“扰动”现有的系统。

那么，问题来了，由元算法生成的新算法还是程序员所创造的吗？科学家开始认识到，真正的新事物可脱胎于旧事物的组合，而整体的行为远比构成它的部分复杂。目前科学界对涌现理论<sup>[1]</sup>较为推崇，它是对还原论观点的纠正。在还原论观点中，一切都可以还原成简单的原子和基本规律。但我们发现意识和水的湿润特性都应归为涌现现象，因为一个单一的水分子不会有湿润的特性，只有一组水分子在某一时刻作为整体才具有湿润性。类似地，一个神经元没有意识，但许多神经元在一起构成神经网络就可以产生意识。学术上有一个很有趣的推断：基于涌现现象的观点，时间的概念并不是绝对的，它的出现是人类对宇宙认知不足的产物。

我们应该把那些新的复杂算法产出的结果看作“涌现现象”。这些结果都是创造它们的规则的产物，但这些结果的整体行为往往大于组成它们的各部分的总和。一些艺术家，尤其是小说家，一旦开始创作，整个过程就好像自己拥有了生命一般。威廉·戈尔丁（William Golding）觉得他的故事似乎变得不受自己支配了：“作者变成了观众，一同感受，或惊或喜。”如果要证明洛夫莱斯的预言是错的，那么关键在于证实程序员和代码之间也存在类似的关系吗？

有一种猛烈抨击人工智能创造力的声音是这样的：它无法反思自己的输出，无法判断其是好是坏，是值得分享还是直接删除。但是，这种自我反思的能力被证明是可能实现的。人们可以创建具有对抗性的算法，来判断一件艺术品是过于因循守旧，还是偏离了我们所认为的艺术的边界。那么，为什么我仍然认为，即使是这些令人惊叹的新工具，也远远无法与人类的创造力相媲美呢？

到目前为止，机器所有的创造力都是由人类的思维和智慧来启动和驱动的，我们还没有发现机器被强制去表达自己。除了我们让它们去做的事情，它们似乎没有什么别的可表达。就像曲艺中的双簧，它们是在台前表演假动作的人，为在台后渴望表达自己的我们提供了喉舌。这种创造性的冲动是我们对自由意志信念的表达。我们可以像机器一样过着日复一日、一成不变的生活，也可以突然做出选择，停下来打破常规，创造新事物。我们的创造力与我们的自由意志密切相关，这似乎是不可能自动化的。创建拥有自由意志的程序与自由意志的含义本就是一组悖论。最终我们可能会发问，我们所拥有的自由意志是否只是一种假象，只是用来掩盖我们自身潜在算法复杂性的“遮羞布”。

目前，人类创建具有创造力的算法的动力，在很大程度上不是由于扩大艺术创作的欲望，而是为了增加商人们在银行的存款。关于人工智能有大量的炒作，有太多的项目打着人工智能的旗号，但其实它们只不过是统计学或数据科学。就像在世纪之交时所有的商业公司都希望成功地在公司名称的末尾加上“.com”一样，现如今在公司名称中加上“AI”或“Deep”，正是这些商业公司赶时髦所利用的标签而已。

商人们希望让人们相信人工智能太伟大了，几乎无所不能，它可以自己写文章、作曲、绘画。这一切都是为了让投资人相信，如果他们进行投资，人工智能将改变他们的业务。但是，当抛开炒作，你会发现驱动这场革命的仍然是人类的思维和智慧。

我们对创造力的痴迷其实是很有趣的。“创造性”一词的意思是“具有价值的新奇事物”，这实际上是20世纪资本主义对“创造性”一词的典型诠释。它源于20世纪40年代广告业高管亚历克斯·奥斯本（Alex Osborn）写的“心灵鸡汤”类书籍。像《你的创造力》和《头脑风暴》这样的书籍，都希望引导人在个人层面和组织中实现创造力。但在商业化“创造性”之前，创造性活动的目的在于捕捉人类试图理解世界存在的意义。

我们可以继续像机器一样在世界上行尸走肉般活着，或者我们可以选择打破规则的限制去理解我们存在的意义。正如心理学家卡尔·罗杰斯在他的文章《走向创造力理论》中所表达的那样，“它是扩大、延展、进化、成熟的冲动，是一种表达和激活躯体所有能力的倾向，这种能力的激活增强了躯体或‘自我’。”正因如此，拥有创造力的是人类而不是机器。今天的人工智能与人类的创造力相比还有一段距离，但它在促使我们更具创造力方面发挥了巨大的作用。不可思议的是，它给了我们在日常生活中经常缺失的点燃创造性的“火镰”，最终它可能会帮助人类减少机械重复的行为。

归根结底，我认为罗杰斯分析中的“自我”这个词是关键所在。我认为，人类的创造力和意识是密不可分的。如果没有意识的概念，我们就无法理解什么是创造力，我们为什么会有创造力。虽然不可能有方法去证实，但我怀疑这两者是同时出现的。伴随着我们精神世界的构建与实现，人类渴望了解自己，并将所得分享给那些无法进行“自我”创造的人。对于巴西作家保罗·科埃略（Paulo Coelho）来说，这种渴望是人性的一部分：写作意味着分享。他想要分享的东西是人类的一部分——想

法、观念、观点。杰克逊·波洛克说：“绘画即探索自我。每个优秀的艺术家都会画出他自己。”意识的问题之一就是，任何人都不可能真正理解其他人：你的疼痛和我的一样吗？你在极度快乐的时刻所感受到的狂喜和我的是同样的感觉吗？这是科学永远无法回答的问题。对于情绪而言，一则故事或一幅绘画要比一台试图扫描我们情绪状态的核磁共振扫描仪更好。对于创造性艺术而言，音乐与文学可能是描绘探索我们作为有意识、有情感的人类到底意味着什么的最好的“画布”。

小说家艾略特写道：“无论是画家、诗人，还是文学家，我们对这些艺术家最大的感激之处是他们将我们的同情心延伸了出去……艺术是最接近生命本质的东西，这是一种放大的体验，是一种超越个人命运与他人接触的方式。”

艺术在调解个人与群体的关系中所扮演的政治性角色也是关键的，它往往是关于改变现状的愿望：打破人性，打破当前的游戏规则；为我们的人类同胞创造更好的，或者仅仅是不同的东西。这无疑是乔治·奥威尔（George Orwell）的动机之一：“当坐下来写一本书时，我不会对自己说，我要制作一件艺术品。我写这篇文章是因为我想要揭露一些谎言，表达一些我想引起大家注意的事实，而最初我关心的是能促成举办一个听证会。”对于扎迪·史密斯（Zadie Smith）来说，她讲故事有一种政治动机：“写作是我表达的方式，它可以帮助消除我们可能会误入歧途的各种方式。”

为什么人们会成为这些艺术作品的欣赏者？也许欣赏艺术作品在一定程度上就是一种创造性的行为，欣赏者参与其中。通常需要一些创造力才能与许多艺术作品产生欣赏的联结，这些作品特意为观众、读者、听众留下空间来发挥他们自己的想象。朦胧是艺术创作的重要组成部分，因为这是欣赏者可以发挥创造力的地方。

有人认为，我们的生活就是一种创造性的行为。莎士比亚是最早意识到这一点的人之一，他在其著名诗作As You Like It中这样写道：

### 【原文】

All the world's a stage

And all the men and women merely players

They have their exits and their entrances

And one man in his time plays many parts

### 【译文】

世界是个大舞台

男男女女演出来

都有谢幕与出场

一生扮演几多角

美国心理学家杰罗姆·布鲁纳（Jerome Bruner）认为：“‘自我’可能是我们创作过的最令人印象深刻也是最复杂的艺术作品。”被我们称为艺术作品的，无论是音乐、绘画，还是诗歌，几乎都是副产品，或者说是我们“自我”创造的一部分。我们又回到了缺乏“自我”的问题上：“自我”的缺位是机器创造力的根本障碍。

创造力与死亡息息相关，也与人类的意义密不可分。许多人在寻找自己存在的意义时，如果发现宗教传说毫无意义，也许会在身后留下一些东西，这些东西将使他们获得“永生”，有时是一幅画、一部小说、一个定理、一个孩子。这些都是企图利用创造力来“欺骗”死亡吗？

也许死亡是我们重视创造力的原因之一。如果柯普真的成功地编写了一种算法，可以大量生成肖邦的玛祖卡舞曲，就好像它让肖邦不朽一样，这会让我们感到开心吗？我不这么认为，反而觉得它会让肖邦创作的作品贬值。这样的它难道不像巴别图书馆吗？包含了一切，却什么都没有。真正重要的是，肖邦的选择。难道国际象棋在某种程度上不是因为计算机的力量被贬低了吗？

也许人类与国际象棋、音乐、数学、绘画的“斗争”，是自身价值实现的一部分来源。许多人认为如果我们彻底解决了死亡的问题，创造出不朽的自己，将会使生命贬值，使活着的每一天都变得毫无意义。在某种程度上讲，我们必死的命运确实很重要——意识到我们必死的命运是意识的代价之一。我的iPhone不可能意识到它将在两年后过时，但是当它意识到时，是否会被什么所驱使，试图留下一些东西来证明自己的存



在呢？

在机器变得有意识之前，我认为它不会仅仅是一个扩展人类创造力的工具。我们知晓怎样才能机器中创造意识吗？在机器中创造意识需要什么？有一些关于人脑神经网络在清醒时和深度睡眠时（我们最无意识的状态）区别的研究，两者的关键区别似乎是反馈的质量不同。大脑在清醒时有意识，活动会从大脑中的一个地方开始，并在整个神经网络中串联，再反馈回原始来源，然后反复重复这个动作序列。这种不停歇的反馈会更新我们的体验。观察处于深度睡眠的大脑，我们只能看局部的兴奋，其没有形成这种反馈的机制。通过机器学习、交互式学习，人工智能具有了某种类似人脑反馈性质的行为特质。这是不是人工智能迈向有意识的第一步，让它最终可以变得真的有意识，然后真正地具有创造性？

但如果机器真的变得有意识，我们怎么去感知呢？它的意识会和我们的吗？我相信在未来，依靠我们人类所有的科学成果，在创造有意识的机器的道路上，不会有什么难以逾越的障碍。一旦我们成功了，我希望机器的意识与我们的意识截然不同，我相信它会告诉我们它是什么样的。那时，创造性的艺术将成为我们互相了解的关键。

讲故事与磁共振成像扫描比起来，可能是我们试图理解、掌握像手机一样的智能电子设备更好的途径。这就是为什么到目前为止，在文学创作领域所有所做的努力中，《探索者》带来的感觉是最接近我们期望从有意识的机器中看到的：它是一种试图与人类产生共鸣并理解我们世界的算法。我们开始思考，在未来是否真的有一天当机器变得有意识时，讲故事会是一个重要的工具。当然，机器可能是被强迫讲故事的，而不是像我们一样拥有那种讲故事的冲动。

就像故事是一种强大的政治工具（把人类社会维系在一起），如果机器变得有意识，那么其分享故事的能力可能会把我们从对人工智能有所恐惧的世界中拯救出来（现在科幻题材的作品经常描述未来的机器是多么的恐怖）。小说家伊恩·麦克尤恩（Ian McEwan）在美国发生“9·11”恐怖袭击事件后所表达的，以及他对同理心的重要呼吁，令人难以忘怀：

如果劫机者能够想象、体会乘客的想法和感受，他们就无法继续施暴。一旦施暴者允许自己进入受害者的内心，那么他就很难继续残忍地施暴。想象自己成为自己以外的人会是什么样，这是我们人性的核心。



人性本善，道德之端。

能够通过故事来分享意识世界是我们之所以为人的原因，没有其他物种能做到这样的事。如果机器变得有意识，那么向机器灌输同理心可能会把我们从《终结者》的故事中“拯救”出来。

“Scheherazade-IF”的首席研究员里德尔对该算法生成的一组备选方案中没有选择奇怪的、反人类的选项感到十分震惊。这说明算法可以从人类讲述故事的方式中学习。“最近，我们已经能够证明，接受过故事训练的人工智能除了在最极端的情况下，不会表现出精神失常。因此，计算机的叙事智能可以减轻我们对‘邪恶的人工智能’接管地球的担忧。”里德尔说道。

在异变发生之时，人类的命运将取决于与有意识的机器的互相理解。但正如维特根斯坦所说的，“即使狮子会说话，我们可能也理解不了它”，这同样适用于机器。如果它们变得有意识，那么人类在一开始也不太可能理解它们。最终，会是机器的绘画、音乐、小说、创造性作品，甚至是它们的数学，给予我们机会去破译机器的代码，感受机器的感受。

[1] 在复杂的自适应系统中，凡一个过程的整体行为远比构成它的部分复杂，皆可称为“涌现”。“涌现”指一个系统中个体间预设的简单互动行为造就的无法预知的复杂样态。——译者注