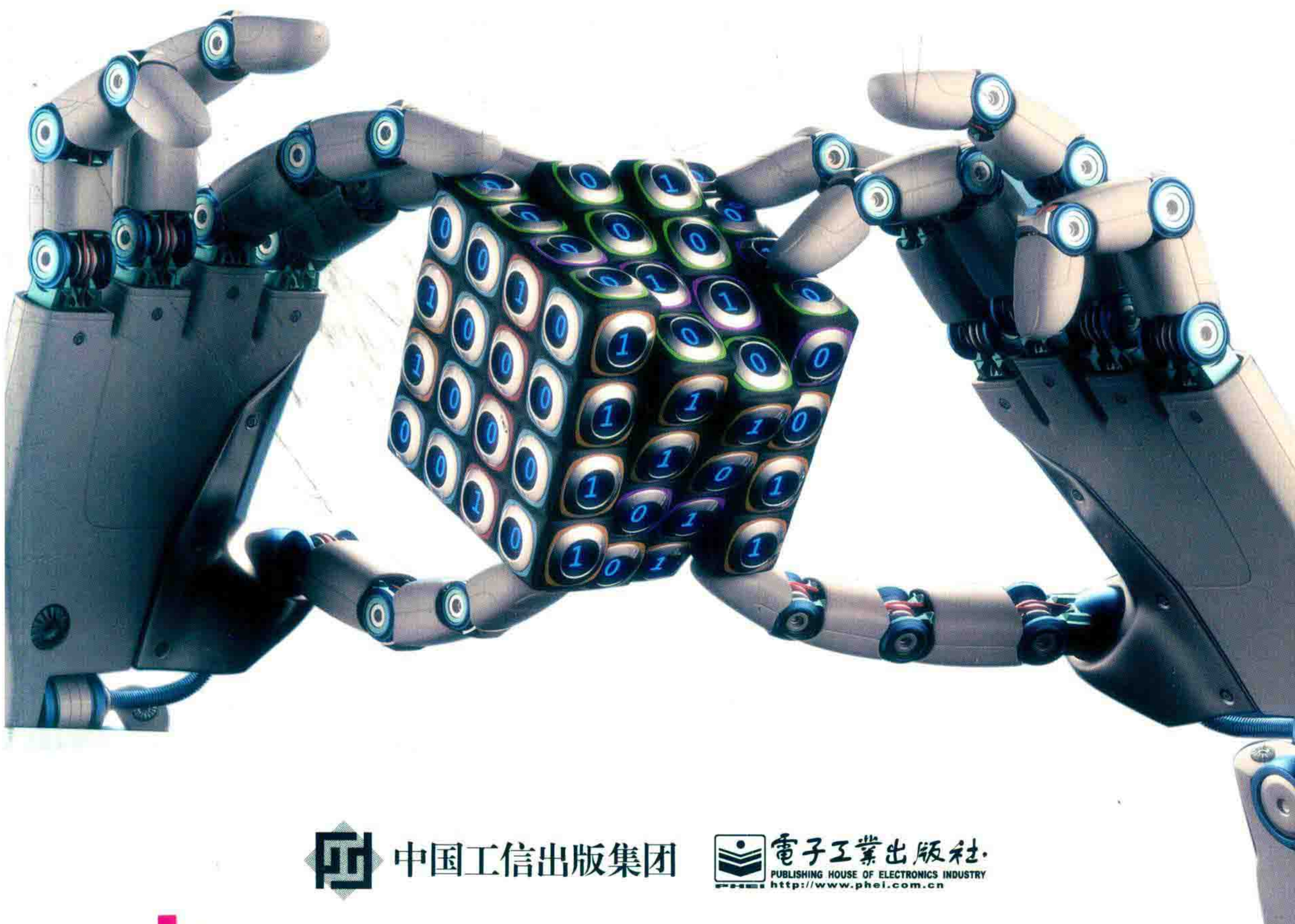


机器学习 VS 复杂系统

许铁 编著 郭瑞东 审校



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

机器学习 VS 复杂系统

许铁 编著 郭瑞东 审校

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书从跨学科视角来看待人工智能这个技术性的学科。围绕用数学模型预测未来这一主题，介绍算法，主要包括现在流行的机器学习和深度学习算法，以及算法要解决问题本身的复杂性。复杂的问题，需要复杂的算法，而算法设计背后的老师正是自然界的复杂性本身。最终，我们上升到自然界解决复杂性最有利的工具，即人类智能本身，让读者从神经科学的角度再次理解人工智能这个大主题，理解神经科学是如何启发人工智能的，而人工智能又如何帮助我们理解人类智能本身。

本书既适合具有高中以上数学知识的一般读者，作为他们了解人工智能和复杂系统领域的科普读物；也适合已经在人工智能领域工作的专业人士，使他们从工程视角之外的更大视角去看待这一领域，获得新的启发。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目（CIP）数据

机器学习 vs 复杂系统 / 许铁编著. —北京：电子工业出版社，2018.8
ISBN 978-7-121-34410-7

I. ①机… II. ①许… III. ①机器学习—普及读物 IV. ①TP181-49

中国版本图书馆 CIP 数据核字(2018)第 120900 号

责任编辑：安 娜

印 刷：三河市华成印务有限公司

装 订：三河市华成印务有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编：100036

开 本：720×1000 1/16 印张：12 字数：192 千字

版 次：2018 年 8 月第 1 版

印 次：2018 年 8 月第 1 次印刷

定 价：59.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010) 88254888，88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：010-51260888-819，faq@phei.com.cn。



序

人工智能是大势所趋，这一主题也成为街头巷尾人们热议的焦点。本书作者的写作初衷是希望能在一个更大的视角下看待人工智能这个技术性的学科。虽然人工智能的技术更多用于工业界，但是其产生的根源，却与人类用数学模型探索世界和自己大脑本身的历程密不可分，因此，本书希望在这个大背景下，从高度跨学科的视角切入人工智能这个大主题。

人人都想预测未来，但是无论科技多么发达，预测未来依然是困难的，比如股市、自然灾害、一个月后的天气预测等，都很难做到精准。究其根本原因，在于无处不在的复杂性。复杂系统这门学科可以帮助我们理解复杂性产生的根源。

大数据时代，数据已成为人们最大限度减少这种不可预测灾难的工具。然而过度信赖数据，往往会让我们陷入一种新的迷信，或者埋没在噪声里。

算法，主要包括现在流行的机器学习和深度学习算法，可以帮助我们最大程度地从数据里提取信息、剔除噪声，这也成为目前人工智能的基础。从另一个角度看，即从人类智能本身来看待这个问题，会让我们对机器学

习和深度学习的本质有着更新的理解，并发现两者内在的相似性，同时看到机器的算法和人脑的算法这两种相似性的根源，即世界本身的复杂性。

本书内容涉及机器学习基本方法、非线性动力学、复杂系统、随机过程、神经网络等正在深刻改变世界的学科，讲述人工智能和复杂系统的基本原理，以及它们是如何影响我们日常生活的方方面面的：小到微信里的语音识别，大到我们到底能不能预测经济危机或金融危机。

我们希望用化繁为简的笔墨，带领读者逐步理解这个大视角下的核心概念，掌握进入更多细分领域的语言和思维体系。我们的使命不是涵盖这一宏大领域的所有知识点，而是希望引发读者的兴趣，去寻找和了解这一激动人心的领域的更多进展。

本书既适合具有高中以上数学知识的一般读者，可作为他们了解人工智能和复杂系统领域的科普读物；也适合利用业余时间充电，正在自学人工智能的工程师，作为他们搭建框架和进入专业领域的开胃菜；还适合已经在人工智能领域从业的专业人士，使他们从工程视角以外的更大视角去看待这一领域，获得新的启发。

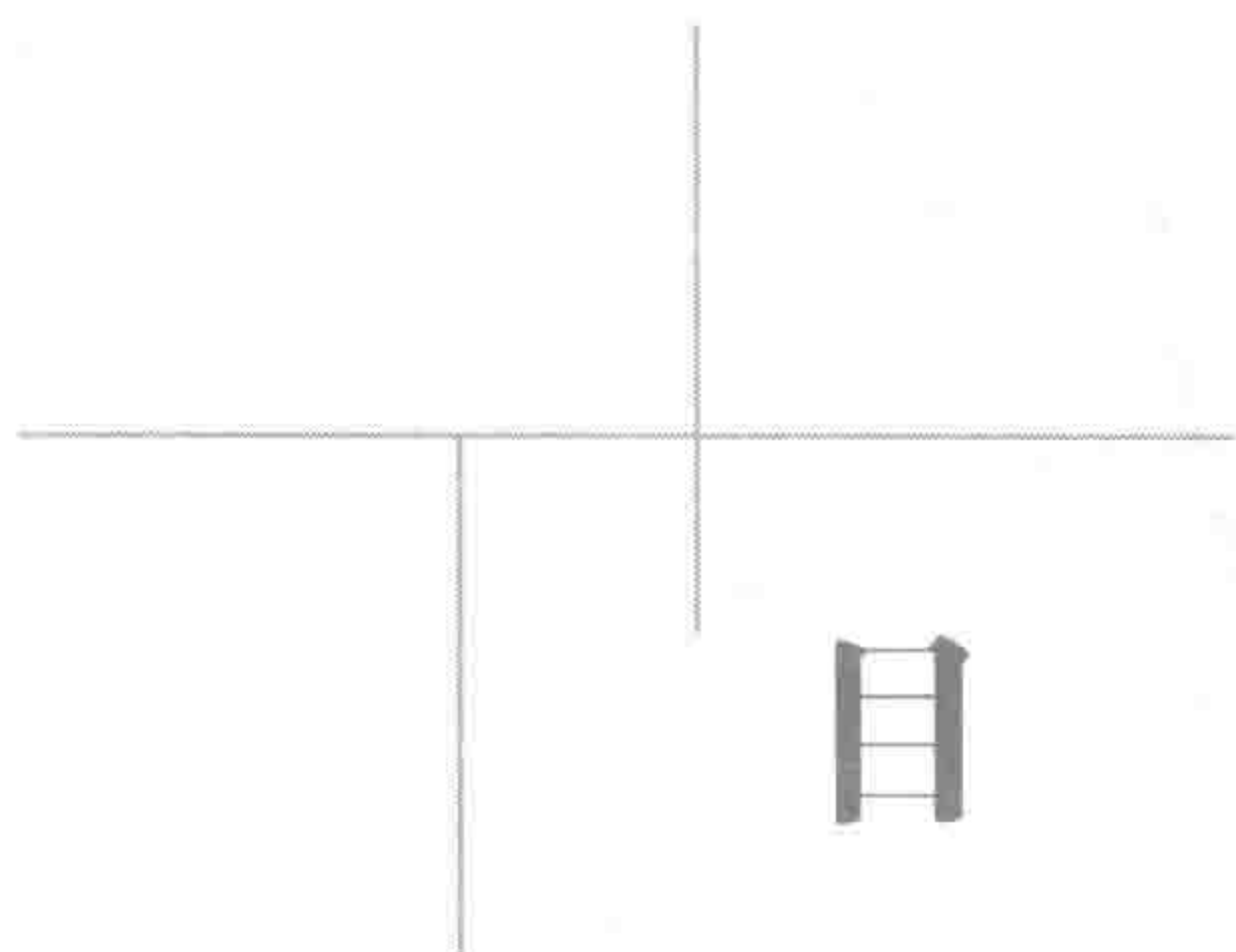
读者服务

轻松注册成为博文视点社区用户（www.broadview.com.cn），扫码直达本书页面。

- **提交勘误：**您对书中内容的修改意见可在 [提交勘误](#) 处提交，若被采纳，将获赠博文视点社区积分（在您购买电子书时，积分可用来抵扣相应金额）。
- **交流互动：**在页面下方 [读者评论](#) 处留下您的疑问或观点，与我们和其他读者一同学习交流。

页面入口：<http://www.broadview.com.cn/34410>





目 录

第一部分 复杂性

1	复杂系统.....	2
2	用复杂网络看世界经济（阅读难度★）	7
3	风险管理策略之复杂科学视角	11
4	从物理角度看复杂	24

第二部分 机器学习

5	白话机器学习（阅读难度★）	46
6	浅谈贝叶斯分析	53
7	简单贝叶斯分类器（阅读难度★）	57
8	决策树方法（阅读难度★★）	60
9	感知机：神经网络的基础（阅读难度★★★）	64
10	降维：应对复杂的通用武器（阅读难度★）	67

第三部分 神经网络

11 神经网络不神秘 74

12 CNN 的几个关键词（阅读难度★★★） 80

13 时间序列与 RNN 91

14 会遗忘的神经网络（阅读难度★★★） 96

15 跟着 AlphaGo 理解深度强化学习框架（阅读难度★★★） 100

16 从阿尔法元看强化学习的更广阔潜力 107

第四部分 宇宙间最复杂的就是我们的

大脑

17 深层视觉信息的编码机制（阅读难度★） 114

18 大脑的自由能假说——兼论认知科学与机器学习（阅读难度★★） 121

19 大脑中的支持向量机（阅读难度★★★） 126

20 机器学习是如何巧妙理解我们大脑的工作原理的
（阅读难度★★） 133

21 大脑经济学（阅读难度★） 140

22 人工智能 vs 人类智能（阅读难度★★） 149

第五部分 人工智能应用谈

23 人工智能会取代艺术家的工作吗 156

24 机器学习预测心理疾病 159

25 人机协作决策的两种方式 164

26 小数据机器学习 166

27 用深度学习玩图像的七重关卡 170

28 深度学习助力基因科技 174

29 机器学习对战复杂系统 176

第一部分 复杂性

1

复杂系统

何为复杂（阅读难度★）

复杂系统用来阐述模式的产生。现实生活中模式无处不在，例如，同类型的商店往往比邻而居；分久必合，合久必分，热恋过后不是分手就是回归平淡。甚至那些我们不认为是模式的，比如生命过程本身，其实都可以看作自然界中模式涌现的过程。这些模式往往可以归结为组成系统的个体通过简单相互作用达到某种非同寻常的集体的现象，复杂系统用动力学的思想阐述了这一过程的发生。

（1）系统可以被分解为比自身简单许多的个体。市场→人，生物→细胞。但是个体和系统之间却呈现了完全不同的属性，用完全不同的特征描述。比如，人具有的生物属性是吃喝拉撒，而由很多人组成的社会系统，就具有了国家、政府、交易市场等这些特征。

（2）个体之间的相互作用反过来受到系统制约。比如，人与人交换物品，需要依靠共同的媒介——钱，而依据的标准是大量交易行为中涌现的价格。这些制约因素往往是在个体的相互作用中自发涌现的。这些制约可以看作一种宏观秩序或者组织，使得相隔遥远的个体通过它联系起来。

一些非常简单的个体之间的相互作用规则就可以诞生出非常复杂的宏观范式,这一现象最早受到物理学的启发。一个里程碑式的模型被称为 ising 模型,这个模型利用不同条件下简单晶体的演变阐述了最复杂的物理问题——相变,在后面会继续介绍。

(3) 系统与外界环境的作用可以用信息的流入和流出描述。复杂系统在外界信息发生变化时的反应是描述复杂系统属性的根本,很多能够稳定存在的复杂系统都具有相对外界环境自发调整自己结构的能力。

最典型的复杂系统的例子就是股市,大家都想预测股票价格,但是很难预测,这是因为股票价格本身即是大量交易者买卖之间涌现的一种宏观秩序,它与公司的业绩并不明显相关,公司业绩对股市的影响更像是进入交易市场这个复杂系统的一个外部信息,外部信息影响交易者的心理但不决定他们的行为,最终股票的价格是由交易者间的相互作用(博弈)决定的。如果要研究股市,更多的是研究如何根据交易者所透露的蛛丝马迹推测其可能的行为。即使你能够跟踪每一位交易者,也由于上述几个原因,股市依然很难预测。

金融市场的不确定性的根本来自于复杂系统,这使得我们常常低估风险。比如次贷危机,我们依据的假设很多时候是把市场看作独立作用的部分,每个部分的风险是独立的。事实上市场的每一部分并没有独立,正是它们之间的互相关联导致次贷危机的。

为何复杂(阅读难度★★)

复杂系统的复杂,是由哪些因素影响造成的呢?

作用(关联)

不是单体的特性,而是单体是如何相互关联形成组织的。因为这类系统共同的特点是长程关联。关联往往导致 $1+1>2$ 或 $1+1<2$, 或称为非线性。市场中出现的价格是受网络相互作用导致的,我们都受到邻居的影响(Herding Effect),相互作用非常重要。比如,神经元就是通过相互作用构成神经网络来处理信号的。

相互作用使得物理系统无法轻松地由整体拆成部分,非线性即其根源,那么还有

什么方法可以解决这个问题？

一旦这种相互作用的维度增加，就会产生一个新的现象，即混沌。混沌说的是由于系统内自由维度的增加，系统的动力学属性不再归于闭合轨道，而是开放或成为不可预测的轨迹，初始条件的轻微变化在未来的影响远未可知，想想真是脊背发凉。

相互作用导致协同效应。两个人在一起可以是 $1+1>2$ ，也可以是 $1+1<2$ ，但基本不会是 $1+1=2$ ，前两者可以看作非线性的体现。比如为什么会有公司，那一定是某种合作导致 $1+1$ 大于 2 的效应使得公司可以产生。

反馈

复杂系统多用于描述一个系统的时间变化过程，比如市场价格的波动、神经网络随时间的活动等。研究这个时间变化过程，往往要考虑此刻的结果对下一刻系统输出的影响。股市的反身性就是反馈的一种。

反馈分为正反馈和负反馈两种，负反馈导致定点平衡态，正反馈导致不稳定性，如雪崩、股市崩盘等。在所有复杂系统中，都有正反馈和负反馈两种状态。反馈带有回路的概念。一个单元通过相互作用传递给另一个单元，反过来另一个单元又可以把信息传递回来。反馈往往是指此刻活动对下一刻活动的影响，比如市场价格。市场价格永远围绕均衡波动，价格高，导致市场买的人变少，买的人少后又导致价格降低，这是典型的负反馈。负反馈把系统维持在稳定位置。 $dx=-x$ ，这是负反馈。

相变

这是复杂系统的第三个重要特质，而且是组织形成的核心。当系统主导反馈的性质发生变化时，则经历一个相变。

相变在自然界和社会中无处不在，自然界中的相变既包括冰和水之间的转化，也包括磁铁从一种相到另一种相的变化。物理中相变的典型例子是磁铁的 ising 模型。

磁铁这个东西，并非总具有磁性。那么具有磁性和不具有磁性的磁铁有什么区别呢？

磁铁有两个相，一个是组织成分均匀一致（有序）的状态，另一个是无序而混乱的状态。虽然它们都是由铁原子构成的，但铁原子只有在有序排序时才会产生磁性。

当需要对外发挥一种作用时，就需要齐心合力，而无序的铁原子使得每个磁针的磁性相互抵销了。

这里就建立了相的概念。而相变，就是通过外部变量使得整个系统从一个相到达另一个相的过程。相变理论是复杂系统研究的重要对象，我们知道，磁铁有的有极性，有的没有极性。研究磁铁极性变化的模型被称为 ising 模型，说的是 paramagnetic（无磁性）到 ferromagnetic（有磁性）的变化。这里影响一个系统相变的主要因素有两个：一个是熵（无序性，系统信息的缺失），另一个是某种趋同的效应。

在铁磁物质里，每一个原子都有极性，平行排列的原子具有指向相同方向的趋势，而熵无序时则破坏了这种效应，两种力量互相争夺。在较高温度下，熵的作用占主导；而较低温度下，有序的趋同的力量占主导。在某个温度下，磁体的原子从无序的状态过渡到完全有序的状态。在完全有序的状态下，整个磁体就显现出对外的磁性。

这里，我们可以控制的外部变量就是温度。温度越高，熵越大。当温度为 0 时，系统自由能最小的状态是一致有序的态，温度升高，无序的态的自由能逐渐减少，直到某个点，成为更有优势的态。类似的还有水到冰的相变，也是在某个温度上，无序和有序交替，这称为临界。所谓临界，就是相变时候的状态，因为这个时候最为特别。临界点上的系统属性特别复杂，统计上我们经常会看到肥尾分布或类似肥尾的分布，这样的分布无处不在，比如股市价格波动、工资分布（帕累托）等。

临界极为重要，为什么？

因为系统在临界点上的属性特别复杂、丰富且有趣，而且，更重要的是，大部分和我们息息相关的系统事实上都在某种程度上处于临界态（或靠近临界态），包括大部分的生物系统和经济系统。

更多关于临界态的内容，可以参考《大自然如何工作》这本书，知乎上的大 V 傅渥成的回答。

“此外，所谓涌现，是在刚才讲到的作用、反馈以及自组织临界基础上得到的，系统从微观到宏观，性质而非数量上‘从零到一’式的改变。”

最简单的例子是路，所谓人走得多了就成了路，森林中交错的小径是大量人走过所涌现出的一种现象。涌现性和相变点也有千丝万缕的联系，有兴趣的读者可以关注自组织临界（Self Organised Criticality）的理论，去查看更多这个领域的知识。

复杂系统的元素很多，而且元素之间均有相互作用，最好的刻画方法就是复杂网络。

2

用复杂网络看世界经济

(阅读难度★)

图 2.1 描述的是一个叫作产业森林的概念。产业森林描述了一个由种类繁多的农业和工业产品组成的关系网。每一个产业即网络里面的节点,就像产业森林里的一棵树,例如图上标出的谷物、采矿业、电子产品等。这里说的是,如果把这些产业描述成复杂网络的一个点,那么我们会对这件事有一个全新的理解。

如何把产业和复杂网络联系起来呢?首先,要做出一个网络结构,我们必须有距离和连接的概念。事实上,不同产业之间不是孤立的,这里所说的产业之间的相似度,其实就是构成它们的技术或生产资料要素的相似性。我们可以规定超过一定相似度阈值的两个产业就可以被联系起来,这样,即可得到不同节点两两相连的矩阵,画出来就能得到产业的复杂网络图。

森林中心是高级工业品,森林边缘是农产品和原材料等。森林中心树木密集,而森林边缘树木稀疏。值得注意的是,树木密集的地方树木间也更盘根错节(连接密集),而树木稀疏的地方连接也越来越少。在森林的边缘,有些树木基本就是孤立的,比如说奶牛生产。这个图颇像城市的交通图,市中心路网交错,而到了郊区的小村,基本就到了路的尽头。



图 2.1

这个网络如何理解？

第一，每个产业间的连接刻画出了两种产品间在生产过程中的联系，外围的产品是内侧产品的原材料，当然内侧产品也可以给外围产品提供服务。第二，连线的颜色对应两种产品的相似度。何为相似度？这个概念可用来刻画生产它们时需要的生产要素的相似性（如劳动、技术、资本……）。例如，梨和苹果的相似度就很高，如果一个地方适合栽种梨树，那么它也往往适合栽种苹果。因为梨和苹果的生长要素差不多，同样的气候和土壤，都需要大量劳动力投入。节点的颜色被用以描述产业的大类，比如资本密集型都用蓝色（图 2.2 中的⑦）表示，劳动密集型都用蓝色（图 2.2 中的⑧）表示。节点的大小代表年交易量，如图 2.2 所示。

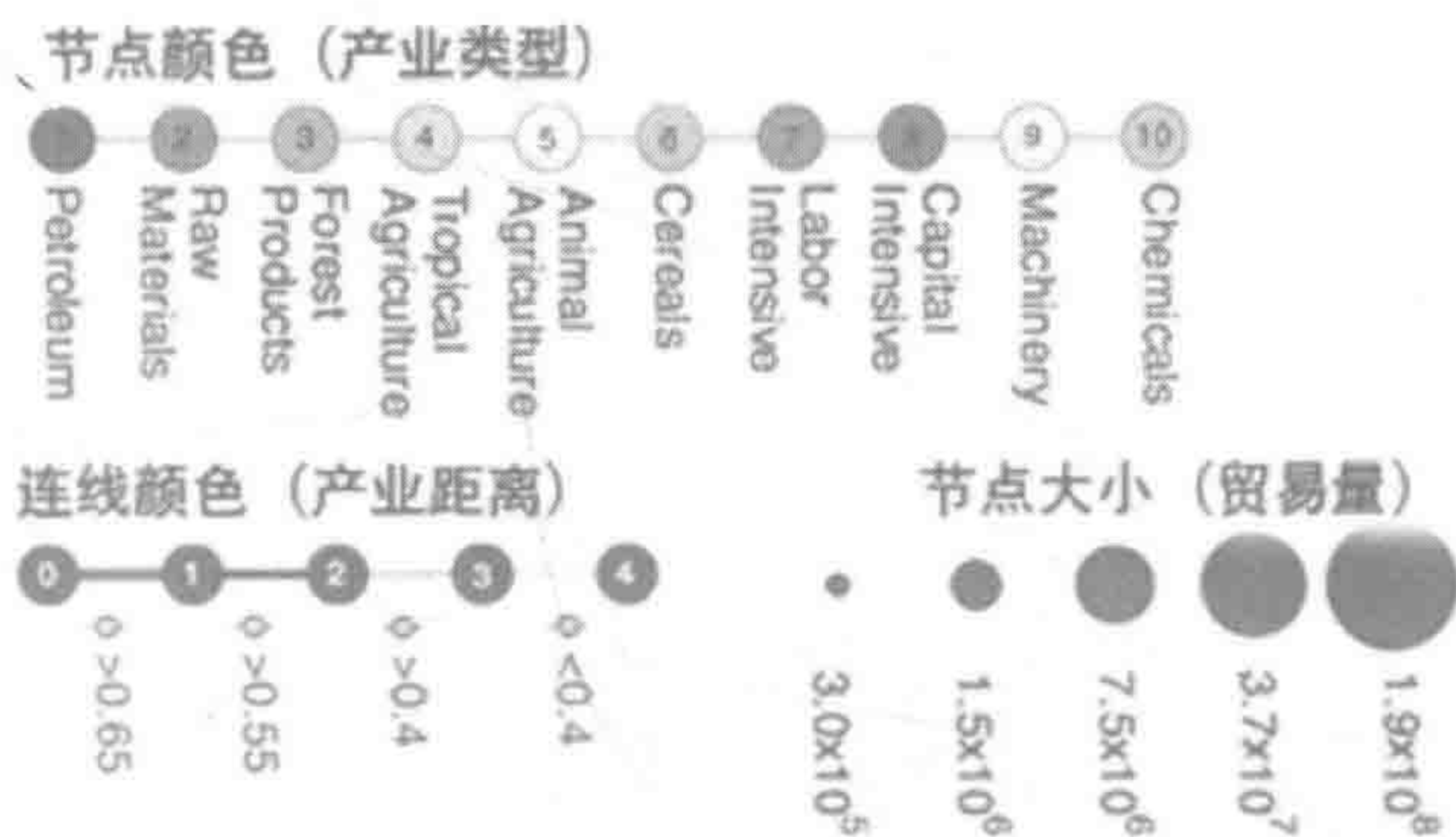


图 2.2

重新回去看产业森林的图，就会立刻明白，森林的边缘树木稀疏而中心树木繁密（树木的间距即产品的相似度，相似度小则间距大，注意树木的间距是用颜色表明的，原图见下载资源）。边缘的树木果实稀少（小圆，贸易量小），而中心的树木硕果累累（大圆，贸易量大），石油产业除外。

换句话说，这张图对全球的所有产业都做了一个聚类，越靠近中心的产业附加值越高，边缘的则与之相反。

这样一个网络结构究竟有着何种重要意义呢？传统理论认为这样的一个结构图没有什么用，因为条条大道通罗马，只要经济在积累，总有机会到达果实丰硕的森林中心。

而复杂网络理论则告诉我们一个完全不同的故事。首先，一个企业往往用它所占据的较为优势的产业来表示。一个企业的发展，被称作从森林的某些位置向其他位置跃迁的过程。可以把企业比喻成产业森林里跳越的猴子，它可以从一棵树跳向另一棵树（产业升级）。当然猴子的目标都是朝着果实丰硕的中心去的。以往的文章往往重视研究猴子，如果一个猴子长期无法到达中心，一般认为是猴子的问题。一个预先的假设是猴子只要跳得足够好，总能从一棵树跳到另一棵树并到达森林的中心。

下面我们从森林的结构来阐述为什么有的猴子能够到达森林的中心而其他猴子不行，因为森林的结构是很重要的制约因素。

网络的性质至关重要，因为有些网络结构，任何一个猴子经过足够长的时间都可以到达森林的中心；而对于另一些网络，猴子能否到达中心，则完全取决于它初始所在的区域，在某些区域，即便它有三头六臂也无法到达森林的中心。

前面提到过，森林的边缘树木如同独立的孤岛，而中心的树木密集（产品相似度大），如同大陆。猴子的跳跃能力是有限的，猴子从森林外侧向中心跳跃不是一件容易的事情。当猴子所在树的周围树木过少且树木间距过大时，猴子就无法跳到下一棵树上。森林中心的猴子可以很轻松地在树木之间跃迁并摘取丰美的果实，而边缘的猴子则没什么选择。

另外一个发人深省的现象是猴子能否进入森林的中心和它所在的初始位置以及它选择的跳跃方向十分相关。如果它开始所在的位置恰好树木间距不太大，而且存在能够到达森林中心的道路，那么它将很有机会进入中心，反之则很难。即便如此，如果一只猴子没有选择好方向，不小心跳到了独木上，四周没有其他树，那么它被困在那个地方的概率就很大。这里其实也有不少例子，当然只考虑森林的结构肯定是偏颇的，但森林的结构的确是很多例子背后的隐藏变量。

图 2.3 和图 2.4 是产业森林的局部截图。从图 2.3 可以看到，劳动密集型工业（汽车制造）和电子产品工业是一条由森林边缘通往中心的捷径，品种繁多，相距很近。从图 2.4 中我们可以看到热带作物（咖啡和可可）一直处在森林的边缘，与森林中心差距较远（仅橡胶与工业品联系较密切），各种热带作物间也距离较远。



图 2.3

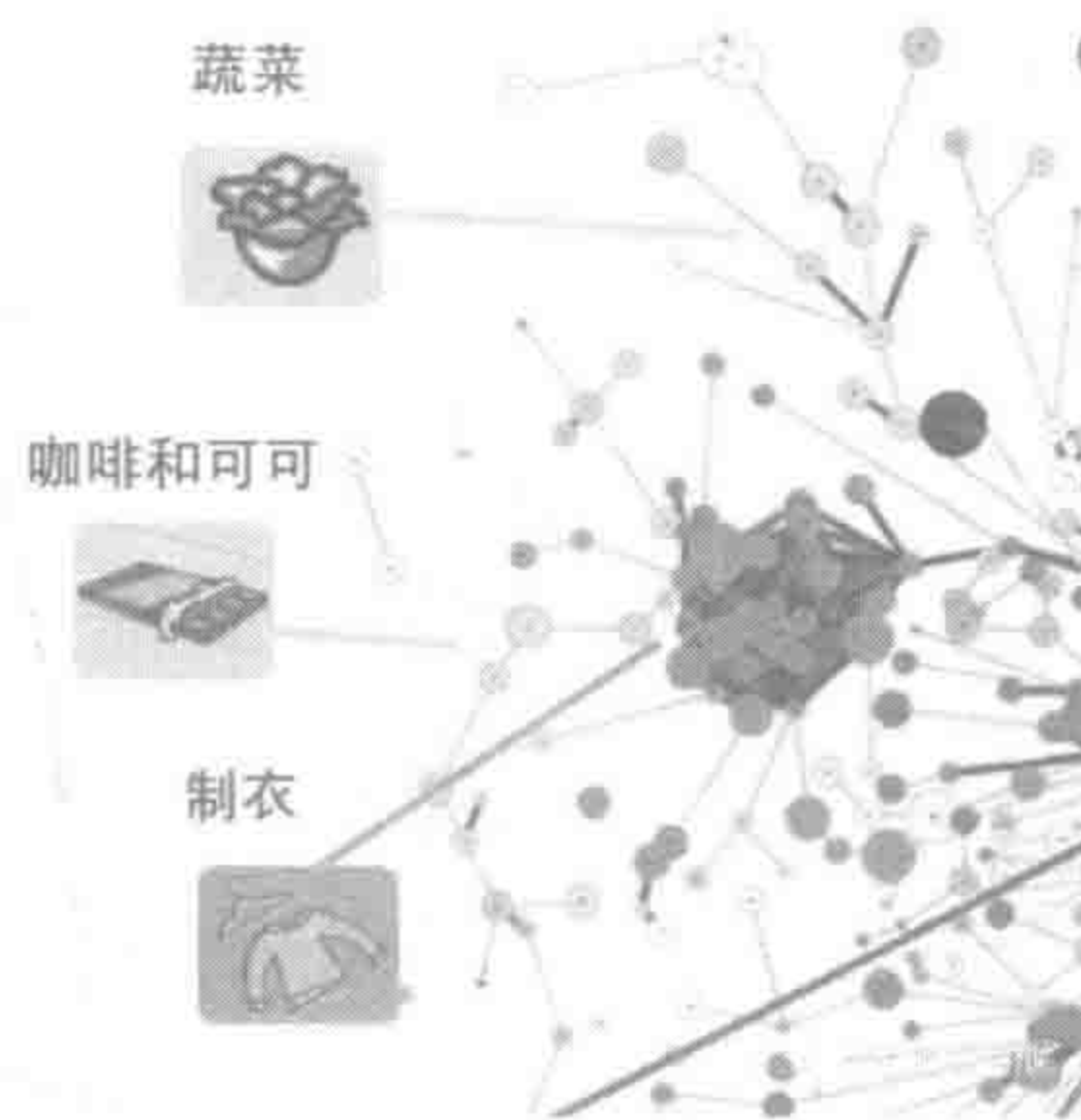


图 2.4