

智能时代, 吴军

机械论已经完全无法对未来进行预测。 香农，这位不世出的天才，则通过借用热力学中“熵”的概念，引入“信息熵”，用信息论将世界的不确定性与信息联系在了一起。这个建立在不确定性上的理论，正是今天人类研究大数据与机器智能的基石。

吴军, 智能时代. Kindle Edition. loc. 51-53. Accessed: 2/1/2017

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

吴军, 智能时代. Kindle Edition. loc. 54-55. Accessed: 2/1/2017

互联网上的任何内容，比如文字、图片和视频都是数据；医院里包括医学影像在内的所有档案也是数据；公司和工厂里的各种设计图纸也是数据；出土文物上的文字、图示，甚至它们的尺寸、材料，也都是数据；甚至宇宙在形成过程中也留下了许多数据，比如宇宙中的基本粒子数量

吴军, 智能时代. Kindle Edition. loc. 153-155. Accessed: 2/1/2017

数据中隐藏的信息和知识是客观存在的，但是只有具有相关领域专业知识的人才能将它们挖掘出来

吴军, 智能时代. Kindle Edition. loc. 195-196. Accessed: 2/1/2017

约翰内斯·开普勒(Johannes Kepler, 1571—1630)。开普勒在所有一流的天文学家家中，资质较差，一生中犯了无数低级的错误。但是他有两样别人没有的东西，第一是从他的老师第谷(Tycho Brahe, 1546—1601)手中继承的大量的、在当时最精确的观测数据，第二是运气。开普勒很幸运地发现了行星围绕太阳运转的轨道实际上是椭圆形的，这样不需要用多个小圆套在大圆之上，而只要用一个椭圆就能将星体运动规律描述清楚

吴军, 智能时代. Kindle Edition. loc. 274-278. Accessed: 2/1/2017

。根据王进喜穿的厚棉袄和戴的大皮帽，可以断定油田定是在中国极北的地区，日本人估计油田应该在哈尔滨和齐齐哈尔之间。其次从背景中井架的密度，大致可以估算出油田的产量。最后从王进喜握手柄的方式，大致能推算出油井的直径。

吴军, 智能时代. Kindle Edition. loc. 304-307. Accessed: 2/1/2017

到了2007年，帕特尔突然在全世界声名鹊起，因为他的研究成果被几个工程师开发成了Google的一款产品——Google趋势(Google Trends)

吴军, 智能时代. Kindle Edition. loc. 325-326. Accessed: 2/1/2017

Google的工程师们从4.5亿种关键词的组合中，最终挑出45个重要的检索词条和55个次要词条(归并成12类)作为特征，训练了一个线性回归模型[15]预测2007年和2008年冬季流感传播的趋势和地点，并且将机器预测的结果和疾病控制与预防中心公

吴军, 智能时代. Kindle Edition. loc. 342-345. Accessed: 2/1/2017

Google的工程师们从4.5亿种关键词的组合中，最终挑出45个重要的检索词条和55个次要词条(归并成12类)作为特征，训练了一个线性回归模型[15]预测2007年和2008年冬季流感传播的趋势和地点，并且将机器预测的结果和疾病控制与预防中心公布的数据进行比对，发现准确率高达97%以上。

吴军, 智能时代. Kindle Edition. loc. 342-345. Accessed: 2/1/2017

当然，数据驱动方法要想成功，除了数据量大之外，还要有一个前提，那就是样本必须非常具有代表性

吴军, 智能时代. Kindle Edition. loc. 457-458. Accessed: 2/2/2017

鸟，在东方和西方都有类似的记录，将鸟的羽毛做成翅膀绑在人的胳膊上往下跳，当然实验的结果都可想而知。后来人们把这样的方法论称作“鸟飞派”，也就是看看鸟是怎样飞的，就能模仿鸟造出飞机，而不需要了解空气动力学。

吴军, 智能时代. Kindle Edition. loc. 594-596. Accessed: 2/2/2017

因为狭义相对论就是光速恒定的必然结果。

吴军, 智能时代. Kindle Edition. loc. 1203-1203. Accessed: 2/4/2017

与机械思维是建立在一种确定性的基础上所截然不同的，信息论完全是建立在不确定性基础上，而要想消除这种不确定性，就要引入信息

吴军, 智能时代. Kindle Edition. loc. 1325-1326. Accessed: 2/4/2017

谁掌握了信息，谁就能够获取财富，这就如同在工业时代，谁掌握了资本谁就能获取财富。

吴军, 智能时代. Kindle Edition. loc. 1338-1339. Accessed: 2/4/2017

，用不确定性这种眼光看待世界，再用信息消除不确定性，不仅能够赚钱，而且能够把很多智能型的问题转化成信息处理的问题，

吴军, 智能时代. Kindle Edition. loc. 1340-1341. Accessed: 2/4/2017

，在信息论里用互信息这个概念，实现了对相关性的量化度量。比如通过对大数据文本进行统计就会发现，“央行调整利率”和“股市短期浮动”的互信息很大，这证实了它们之间有非常强的相关性。而“央行调整利率”和“北京机场大量航班晚点”的互信息则接近于零，说明二者没有什么相关性，甚至无关

吴军, 智能时代. Kindle Edition. loc. 1351-1353. Accessed: 2/4/2017

通俗地讲就是信息的传播速率不可能超过信道的容量，

吴军, 智能时代. Kindle Edition. loc. 1375-1375. Accessed: 2/4/2017

Notes: 1) 香农第二定律

当整个搜索行业都意识到点击数据的重要性后，这个市场上的竞争就从技术竞争变成了数据竞争。

吴军, 智能时代. Kindle Edition. loc. 1545-1546. Accessed: 2/4/2017

Google和很多互联网公司之所以能够取得成功，不仅仅是靠技术，靠数据，更是靠采用了大数据时代的方法论，或者说大数据思维

吴军, 智能时代. Kindle Edition. loc. 1567-1568. Accessed: 2/4/2017

亚马逊的优势倒不在于价格便宜，事实上美国实体店和网上的价格差不多，这和中国的电商有很大区别，它的优势是能够有针对性地给用户推荐商品，这占到亚马逊销售额的1/3。为什么亚马逊能够做到这一点而沃尔玛做不到呢？这就

吴军, 智能时代. Kindle Edition. loc. 1740-1742. Accessed: 2/4/2017

新技术+原有产业=新产业

吴军, 智能时代. Kindle Edition. loc. 2048-2048. Accessed: 2/5/2017

现有产业+电=新产业

吴军, 智能时代. Kindle Edition. loc. 2096-2096. Accessed: 2/5/2017

现有产业+摩尔定律=新产业

吴军, 智能时代. Kindle Edition. loc. 2155-2156. Accessed: 2/5/2017

现有产业+大数据=新产业 现有产业+机器智能=新产

吴军, 智能时代. Kindle Edition. loc. 2161-2162. Accessed: 2/5/2017

2010年, Google宣布开发出名为 Google大脑(Google Brain)的深度学习工具。从机器学习理论上讲, 它没有任何突破, 只是把过去的人工神经网络并行地实现了。但是从工程的角度上来讲, 它有非常大的意义。

吴军, 智能时代. Kindle Edition. loc. 2707-2709. Accessed: 2/9/2017

算机系统的设计和高楼设计很大的不同是, 前者事先并不考虑安全的隐患, 而后者在每一

吴军, 智能时代. Kindle Edition. loc. 2787-2788. Accessed: 2/9/2017

算机系统的设计和高楼设计很大的不同是, 前者事先并不考虑安全的隐患, 而后者在每一个环节都要考虑安全的问题, 这就是我们面临的现实。

吴军, 智能时代. Kindle Edition. loc. 2787-2788. Accessed: 2/9/2017

利用大数据(2000万用户产生的36亿行的系统日志)分析来防范黑客攻击, 要比传统的在防火墙设置各种规则的做 法有效5倍。[13]而在工业界, 一些信息安全公司已经开始按照这种思路来设计和研制产品了。

吴军, 智能时代. Kindle Edition. loc. 2794-2796. Accessed: 2/9/2017

当商家知道前者是买了假货也不会吭声的软柿子, 后者是睚眦必报的刺头的时候, 欺软怕硬的行为一定能够给他们带来最多的利益。在利用大数据方面, 个人用户相比商家永远是弱势群体, 一旦他们的秘密被商家知道, 他们的利益就难免受到损害。

吴军, 智能时代. Kindle Edition. loc. 2863-2865. Accessed: 2/9/2017

国很多航空公司在利用个人隐私大发其财。当航空公司发现某个机票的 购票者最近必须旅行, 而且在过去对票价不是很敏感时, 它给出的报价就会比其他人的高很多。尤其当两个城市间 仅此一家航空公司有直飞的航班时, 价格上的差异就更明显。这些航空公司甚至出钱聘请了美国一些著名的大学帮助研究这样利用用户隐私赚钱的方法。据一所世界名校里承接这些项目的团队介绍, 利用对用户行为的分析, 可以让航空公司提高10%左右的销售额。虽然10%的提高听起来不算太多, 但是对净利润率只有0.2%的航空业来说, 这是几十倍的利润的提高。而对于乘客来说,由于只是部分乘客受到伤害, 整体上10%票价的提高意味着他们的额外支出要远比10%多, 实际上多付出的票价可以高达50%。

吴军, 智能时代. Kindle Edition. loc. 2866-2872. Accessed: 2/9/2017

Notes: 1) 美国航空公司利用用户隐私提高利润

美国很多航空公司在利用个人隐私大发其财。当航空公司发现某个机票的 购票者最近必须旅行, 而且在过去对票价不是很敏感时, 它给出的报价就会比其他人的高很多。尤其当两个城市间 仅此一家航空公司有直飞的航班时, 价格上的差异就更明显。这些航空公司甚至出钱聘请了美国一些著名的大学帮助研究这样利用用户隐私赚钱的方法。据一所世界名校里承接这些项目的团队介绍, 利用对用户行为的分析, 可以让航空公司提高10%左右的销售额。虽然10%的提高听起来不算太多, 但是对净利润率只有0.2%的航空业来说, 这是几十倍的利润的提高。而对于乘客来说,由于只是部分乘客受到伤害, 整体上10%票价的提高意味着他们的额外支出要远比10%多, 实际上多付出的票价可以高达50%。

吴军, 智能时代. Kindle Edition. loc. 2866-2872. Accessed: 2/9/2017

它一直把自己定位成一个IT公司, 而不是汽车公司。汽车其实就是承载着特斯拉IT技术的平台, 特斯拉内部 将汽车看成是一个巨大的智能终端, 通过这个智能终端, 特斯拉把它的各种技术服务提供给大家, 同时也参与到消费者的日常生活中, 这和我们前面提到的小米手机有不少相似之处。

吴军, 智能时代. Kindle Edition. loc. 3089-3092. Accessed: 2/11/2017

。汽车这个老行业, 在引入大数据和机器智能之后就脱胎换骨, 变成了一个新的行业。

吴军, 智能时代. Kindle Edition. loc. 3108-3109. Accessed: 2/11/2017