

## 终极算法:机器学习和人工智能如何重塑世界, 佩德罗·多明戈斯

当你用网易云音乐查看推荐歌单的时候, 当你在百度搜索信息的时候, 当你在互联网金融平台申请借款的时候, 甚至在你调戏Siri和小冰的时候, 其实都是其背后的机器学习算法在云端服务器中为你默默服务。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 114-116. Kindle Edition

每个算法都会有输入和输出: 数据输入计算机, 算法会利用数据完成接下来的事, 然后结果就出来了。机器学习则颠倒了这个顺序: 输入数据和想要的结果, 输出的则是算法, 即把数据转换成结果的算法。学习算法能够制作其他算法。通过机器学习, 计算机就会自己编写程序, 就用不到我们了。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 389-392. Kindle Edition

在信息处理这个生态系统中, 学习算法是顶级掠食者。数据库、网络爬虫、索引器等相当于食草动物, 耐心地对无限领域中的数据进行蚕食。统计算法、线上分析处理等则相当于食肉动物。食草动物有必要存在, 因为没有它们, 其他动物无法存活, 但顶级掠食者有更为刺激的生活。数据爬虫就像一头牛, 网页相当于它的草原, 每个网页就是一根草。当网络爬虫进行破坏行动时, 网站的副本就会保存在其硬盘当中。索引器接着做一个页面的列表, 每个词都会出现在页面当中, 这很像一本书后的索引。数据库就像大象, 又大又重, 永远不会被忽略。在这些动物当中, 耐心的野兽飞快运转统计和分析算法, 压缩并进行选择, 将数据变为信息。学习算法将这些信息吞下、消化, 然后将其变成知识。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 428-435. Kindle Edition

企业可以将机器学习应用到企业运作的每个方面, 只要有足够的数据, 只要数据能够从计算机、通信设备以及更廉价、更普适的传感器源源不断地输出。“数据是新型石油”是目前的流行说法, 既然是石油, 提炼石油就是一笔大生意。和其他公司一样, IBM已制定经济增长战略, 为企业提供分析服务。业界将数据看作战略资产: 我有什么数据, 而竞争对手却没有? 我要怎么利用这些数据? 竞争对手有什么数据, 而我却没有?

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 500-504. Kindle Edition

没有数据库的银行无法和有数据库的银行竞争, 没有机器学习的企业也无法跟上使用机器学习的企业。虽然第一家公司的专家写了上千条规则, 预测用户的喜好, 但是第二家公司的算法却能学习数十亿条规则, 一整套规则都可用于每位用户。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 504-506. Kindle Edition

不列颠之战期间, 英国空军阻止了纳粹德国空军的进攻, 尽管后者人数比前者多很多。德国飞行员不明白, 为什么无论走到哪里, 他们总会碰上英国空军。英国有一个秘密武器: 雷达, 可以在德国飞机越境进入英国领空时, 就探测到它们。机器学习就像装了雷达, 能够预知未来。别只是回击对手的行动, 要预测他们的行动, 并先发制人。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 615-619. Kindle Edition

尽管机器学习发展很明显, 但这也仅仅是未来的预告。虽然它有用, 但实际上当今在工业上起作用的学习算法的生成还是受到了很大限制。如果现在实验室的算法能在各领域的前线使用, 比尔·盖茨说机器学习的突破产生的价值将相当于10家微软, 其实这个说法有点保守了。如果这些观点让研究人员真正觉得眼前一片光明, 而且收到效果, 那么机器学习带来的就不仅仅是新的文明时代, 还是地球生命进化的

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 638-642. Kindle Edition

尽管机器学习发展很明显, 但这也仅仅是未来的预告。虽然它有用, 但实际上当今在工业上起作用的学习算法的生成还是受到了很大限制。如果现在实验室的算法能在各领域的前线使用, 比尔·盖茨说机器学习的突破产生的价值将相当于10家微软, 其实这个说法有点保守了。如果这些观点让研究人员真正觉得眼前一片光明, 而且收到效果, 那么机器学习带来的就不仅仅是新的文明时代, 还是地球生命进化的新阶段。 怎样

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 638-642. Kindle Edition

尽管机器学习发展很明显, 但这也仅仅是未来的预告。虽然它有用, 但实际上当今在工业上起作用的学习算法的生成还是受到了很大限制。如果现在实验室的算法能在各领域的前线使用, 比尔·盖茨说机器学习的突破产生的价值将相当于10家微软, 其实这个说法有点保守了。如果这些观点让研究人员真正觉得眼前一片光明, 而且收到效果, 那么机器学习带来的就不仅仅是新的文明时代, 还是地球生命进化的新阶段。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 638-642. Kindle Edition

所有知识, 无论是过去的、现在的还是未来的, 都有可能通过单个通用学习算法来从数据中获得。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 679-681. Kindle Edition

所有这些例子都证明, 大脑自始至终只使用了一种相同的学习算法, 那些负责不同知觉的区域, 区别也仅仅在于与其相连、输入信息的器官(如眼睛、耳朵、鼻子)。反过来, 关联区(大脑的各个皮层)通过与不同的感觉区(各个感觉器官)相连, 来实现其机能, 而执行区则通过连接关联区来实现其机能, 然后输出反馈。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 698-701. Kindle Edition

根据一个统计学流派的观点, 所有形式的学习都是基于一个简单的公式——如我们所知, 就是贝叶斯定理。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 777-778. Kindle Edition

弄明白蛋白质如何折叠成特定形状; 通过DNA来重新构建一系列物种的进化史; 在命题逻辑中证明定理; 利用交易成本来发现市场中的套利机会; 从二维视图中推出三维形状; 将数据压缩到磁盘上; 在政治活动中组成稳定联盟; 在剪切流中模拟湍流; 按照给定回报率找出最安全的投资组合、到达几个城市的捷径、微芯片上元件的最佳布局方案、生态系统中传感器的最佳布局、自旋玻璃门最低的能量状态; 安排好航班、课程、工厂工作; 最优化资源分配、城市交通流、社会福利, 以及提高你的俄罗斯方块分数(最重要的)——这些都是NP完全问题, 意思是, 如果你能有效解决其中的一个问题, 就能有效解决所有NP类问题, 包括相互间的问题。谁会猜到, 这些表面上看起来迥然不同的问题, 会是同一个问题? 如果它们真的是同一个问题, 就可以说一种算法能学会解决所有问题(或更准确地说, 所有能有效解决的例子)。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 796-804. Kindle Edition

弗雷德·贾里尼克(IBM语音研究组的领导)说过一句著名的俏皮话:“每开除一名语言学家, 我的语音识别系统的错误率就降低一个百分点。”

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 870-871. Kindle Edition

像明斯基、乔姆斯基和福多这样的批评家曾经占据上风, 但万幸, 他们的影响力已经逐渐减弱。即便如此, 我们仍需将他们的批评铭记于心, 这样才能到达终极算法这个终点, 原因有两个: 第一, 知识工程师和机器学习算法一样, 遇到许多相同的问题, 虽然他们没有成功, 但学到了许多宝贵的教训; 第二, 学习和知识以异常微妙的形式相互交织, 而我们很快就会发现这一点。遗憾的是, 这两个阵营各说各话。他们讨论不同的主题: 机器学习讨论概率, 而知识工程学讨论逻辑。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 883-887. Kindle Edition

如果你只见过白天鹅, 会觉得看到黑天鹅的概率是0。2008年的金融危机就是一只“黑天鹅”。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 891-893. Kindle Edition

另外一个反对机器学习的观点与以上观点相关，就是我们常听到的——“数据无法代替人类的直觉”。实际上，这句话可以反过来：人类直觉无法代替数据。直觉就是你在不知道事实的情况下依靠的东西，而因为你不常用它，所以直觉非常宝贵。但如果证据摆在你面前，为什么还要拒绝证据？

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 899-901. Kindle Edition

有了机器学习的辅助，经理会变成超级经理，科学家会变成超级科学家，工程师会变成超级工程师。未来属于那些深深懂得如何将自己的独特专长与算法的擅长结合起来的人。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 989-991. Kindle Edition

寻找终极算法更像测定海上的经度，人们一开始认为这太困难，于是放弃了，直到一个孤独的天才解决了这个问题。寻找终极算法更有可能就是一代一代人的任务，就像天主教堂是由一块块石头砌成的一样。找到终极算法的唯一方法就是，早早动身踏上旅途。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1036-1039. Kindle Edition

因为在机器学习领域存在不同思想的学派，主要学派包括符号学派、联结学派、进化学派、贝叶斯学派、类推学派。每个学派都有其核心理念以及其关注的特定问题。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1100-1102. Kindle Edition

对于符号学派来说，所有的信息都可以简化为操作符号，就像数学家那样，为了解方程，会用其他表达式来代替本来的表达式。符号学者明白你不能从零开始学习：除了数据，你还需要一些原始的知识。他们已经弄明白，如何把先前存在的知识并入学习中，如何结合动态的知识来解决新问题。他们的主算法是逆向演绎，逆向演绎致力于弄明白，为了使演绎进展顺利，哪些知识被省略了，然后弄明白是什么让主算法变得越来越综合。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1103-1107. Kindle Edition

*Notes: 1) 符号学派*

联结学派的主算法是反向传播学习算法，该算法将系统的输出与想要的结果相比较，然后连续一层一层地改变神经元之间的连接，目的是为了输出的东西接近想要的东西。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1109-1111. Kindle Edition

*Notes: 1) 联结学派 2) 进化学派*

进化学派认为，所有形式的学习都源于自然选择。如果自然选择造就我们，那么它就可以造就一切，我们要做的，就是在计算机上对它进行模仿。进化主义解决的关键问题是学习结构：不只是像反向传播那样调整参数，它还创造大脑，用来对参数进行微调。进化学派的主算法是基因编程，和自然使有机体交配和进化那样，基因编程也对计算机程序进行配对和提升。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1111-1115. Kindle Edition

进化学派认为，所有形式的学习都源于自然选择。如果自然选择造就我们，那么它就可以造就一切，我们要做的，就是在计算机上对它进行模仿。进化主义解决的关键问题是学习结构：不只是像反向传播那样调整参数，它还创造大脑，用来对参数进行微调。进化学派的主算法是基因编程，和自然使有机体交配和进化那样，基因编程也对计算机程序进行配对和提升。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1111-1115. Kindle Edition

进化学派认为，所有形式的学习都源于自然选择。如果自然选择造就我们，那么它就可以造就一切，我们要做的，就是在计算机上对它进行模仿。进化主义解决的关键问题是学习结构：不只是像反向传播那样调整参数，它还创造大脑，用来对参数进行微调。进化学派的主算法是基因编程，和自然使有机体交配和进化那样，基因编程也对计算机程序进行配对和提升。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1111-1115. Kindle Edition

进化学派认为，所有形式的学习都源于自然选择。如果自然选择造就我们，那么它就可以造就一切，我们要做的，就是在计算机上对它进行模仿。进化主义解决的关键问题是学习结构：不只是像反向传播那样调整参数，它还创造大脑，用来对参数进行微调。进化学派的主算法是基因编程，和自然使有机体交配和进化那样，基因编程也对计算机程序进行配对和提升。贝叶斯学派最关注的问题是不确定性。所有掌握的知识都有不确定性，而且学习知识的过程也是一种不确定的推理形式。那么问题就变成，在不破坏信息的情况下，如何处理嘈杂、不完整甚至自相矛盾的信息。解决的办法就是运用概率推理，而主算法就是贝叶斯定理及其衍生定理。贝叶斯定理告诉我们，如何将新的证据并入我们的信仰中，而概率推理算法尽可能有效地做到这一点。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1111-1118. Kindle Edition

贝叶斯学派最关注的问题是不确定性。所有掌握的知识都有不确定性，而且学习知识的过程也是一种不确定的推理形式。那么问题就变成，在不破坏信息的情况下，如何处理嘈杂、不完整甚至自相矛盾的信息。解决的办法就是运用概率推理，而主算法就是贝叶斯定理及其衍生定理。贝叶斯定理告诉我们，如何将新的证据并入我们的信仰中，而概率推理算法尽可能有效地做到这一点。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1115-1118. Kindle Edition

*Notes: 1) 贝叶斯学派*

贝叶斯学派最关注的问题是不确定性。所有掌握的知识都有不确定性，而且学习知识的过程也是一种不确定的推理形式。那么问题就变成，在不破坏信息的情况下，如何处理嘈杂、不完整甚至自相矛盾的信息。解决的办法就是运用概率推理，而主算法就是贝叶斯定理及其衍生定理。贝叶斯定理告诉我们，如何将新的证据并入我们的信仰中，而概率推理算法尽可能有效地做到

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1115-1118. Kindle Edition

贝叶斯学派最关注的问题是不确定性。所有掌握的知识都有不确定性，而且学习知识的过程也是一种不确定的推理形式。那么问题就变成，在不破坏信息的情况下，如何处理嘈杂、不完整甚至自相矛盾的信息。解决的办法就是运用概率推理，而主算法就是贝叶斯定理及其衍生定理。贝叶斯定理告诉我们，如何将新的证据并入我们的信仰中，而概率推理算法尽可能有效地做到这一点。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1115-1118. Kindle Edition

对于类推学派来说，学习的关键就是要在不同场景中认识到相似性，然后由此推导出其他相似性。如果两个病人有相似的症状，那么也许他们患有相同的疾病。问题的关键是，如何判断两个事物的相似程度。类推学派的主算法是支持向量机，主算法找出要记忆的经历，以及弄明白如何将这些经历结合起来，用来做新的预测。每个

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1118-1122. Kindle Edition

对于类推学派来说，学习的关键就是要在不同场景中认识到相似性，然后由此推导出其他相似性。如果两个病人有相似的症状，那么也许他们患有相同的疾病。问题的关键是，如何判断两个事物的相似程度。类推学派的主算法是支持向量机，主算法找出要记忆的经历，以及弄明白如何将这些经历结合起来，用来做新的预测。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1118-1122. Kindle Edition

机器学习算法和所有科学家一样，类似盲人和大象：有个盲人摸到象鼻，就以为那是蛇；另一个盲人靠着象腿，以为那是树；还有一个盲人摸到象牙，以为那是公牛。我们的目标是，摸清楚每个部位，而不是过早下结论。一旦摸到所有部位，我们就努力拼出整个大象的形象。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1136-1139. Kindle Edition

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1141. Kindle Edition

问题 ( ill-posed problem )：这个问题没有唯一

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1264-1264. Kindle Edition

零售商喜欢一套套的规则，因为他们要决定该囤什么货。通常，他们会用比“分而治之”更为彻底的方法，也就是寻找所有能够准确预测每个购买项的规则。沃尔玛在该领域属先驱，他们早期的发现之一就是，如果你买了纸尿裤，那么很有可能会买啤酒。为什么？对此进行解释的说法之一就是，妈妈让爸爸去超市买纸尿裤，出于情感补偿，爸爸买了一箱啤酒。知道这一点，超市现在会把啤酒放在纸尿裤旁边，这样啤酒就会卖得更好。不找规律，这样的事就不会在沃尔玛发生。“

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1354-1358. Kindle Edition

*Notes: 1) 纸尿裤和啤酒*

伽利略的天才之处在于，无须到外太空亲眼见证，他凭直觉就知道，不受外力影响的物体会一直保持运动。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1395-1396. Kindle Edition

偏差”和“方差”。某座钟如果总是慢一个小时，那么它的偏差会很高，但方差会很低。但如果这座钟走得时快时慢，最后平均下来准点了，那么它的方差会很高，但偏差会很低。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1499-1500. Kindle Edition

*Notes: 1) 偏差和方差*

你可以估算一种学习算法的偏差和方差，方法就是在掌握训练集的随机变量之后，对算法的预测进行对比。如果算法一直出错，那么问题就出在偏差上，而你需要一个更为灵活的学习算法（或者只和原来的不一样即可）。如果出现的错误无模式可循，问题就出在方差上，而你要么尝试一种不那么灵活的学习算法，要么获取更多的数据。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1507-1510. Kindle Edition

苏格拉底是人类。所有人类都会死。所以……

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1527-1529. Kindle Edition

一个细胞就像一台微型计算机，而DNA就是计算机运行的程序：改变DNA，皮肤细胞就会变成神经元细胞，或者小鼠细胞会变成人类细胞。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1570-1571. Kindle Edition

决策树的原理就像玩一个有实例的20问游戏。从“根部”开始，每个节点都会问每个属性的值，然后根据答案，我们沿着这个或另外一个分支继续下去。当到达“树叶”部分时，我们读取预测的概念。从“根部”到“树叶”的每条路线都对应一个规则。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1616-1618. Kindle Edition

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1651. Kindle Edition

把一棵决策树转变成一个规则集很容易：每条从“根部”到“叶子”的路线是一条规则，而且路线不会崩溃。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1680-1681. Kindle Edition

联结学派对符号学派尤其不满。根据他们的观点，你能通过逻辑规则来定义的概念仅仅是冰山一角，其实表面之下还有很多东西是形式推理无法看到的。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1691-1692. Kindle Edition

实际上，目前计算机和人脑相比，主要的限制是能量损耗：人的大脑消耗的能量仅仅相当于一个小灯泡，而沃森消耗的电却能点亮整栋办公楼。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1730-1732. Kindle Edition

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1769. Kindle Edition

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1879. Kindle Edition

该曲线看起来像被拉长的字母S，它有很多叫法，比如逻辑函数、S形函数和S形曲线。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1880-1881. Kindle Edition

*Notes: 1) 逻辑函数*

晶体管的转换曲线，将其输入电压和输出电压联系在一起，也是一条S形曲线。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1882-1883. Kindle Edition

科学中的范式转移、

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1885-1885. Kindle Edition

那么让我们用S形曲线来代替感知器的阶跃函数，  
佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1914-1914. Kindle Edition

这使得反向传播成为联结学派的主算法。反向传播是自然及技术领域非常常见的战略实例：如果你着急爬到山顶，那你就得爬能找到的最陡的坡。这在技术上的术语为“梯度上升”（如果你想爬到山顶）或者梯度下降（如果你想走到山谷）。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1932-1934. Kindle Edition

反向传播是自然及技术领域非常常见的战略实例：如果你着急爬到山顶，那你就得爬能找到的最陡的坡。这在技术上的术语为“梯度上升”（如果你想爬到山顶）或者梯度下降（如果你想走到山谷）。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 1933-1934. Kindle Edition

了，如果把治愈癌症作为目标，我们不必知道癌细胞运转的所有细节，只需要知道如何在不损害正常细胞的

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2021-2022. Kindle Edition

如果把治愈癌症作为目标，我们不必知道癌细胞运转的所有细节，只需要知道如何在不损害正常细胞的情况下，使癌细胞失去繁殖能力，这就足够了。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2021-2023. Kindle Edition

我们不会通过对羽毛进行逆向工程来制造飞机，而飞机也不会拍翅膀。飞机是在气体力学的基础上设计的，所有飞行的物体都必须遵循气体力学原则。我们还是没有理解那些想法的类似原则。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2080-2082. Kindle Edition

联结学派的一些人高调称，反向传播就是终极算法，

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2082-2083. Kindle Edition

联结学派的一些人高调称，反向传播就是终极算法，而我们只需要扩大反向传播的规模。但

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2082-2083. Kindle Edition

联结学派的一些人高调称，反向传播就是终极算法，而我们只需要扩大反向传播的规模。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2082-2083. Kindle Edition

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2268. Kindle Edition

治愈癌症的问题——纯粹受到自然启发的方法要成功可能太无知，即使给定大量的数据也还是如此。原则上，我们可以掌握细胞新陈代谢网络的完整模型，方法就是结合结构研究，利用或者不利用交叉，通过反向传播来进行参数学习，但有太多不利的局部最优陷阱。我们得利用更大块的数据来进行推理，根据需要集合或重新集合这些数据，然后利用逆向演绎来填补空缺。要让这样的目标来引导我们的学习行为：以最优方法诊断癌症，然后找到治疗癌症的最佳药物。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2413-2418. Kindle Edition

对于贝叶斯学派来说，学习“仅仅是”贝叶斯定理的另外一个运用，将所有模型当作假设，将数据作为论据：随着你看到的数据越来越多，有些模型会变得越来越有可能性，而有些则相反，直到理想的模型渐渐突出，成为最终的胜者。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2440-2442. Kindle Edition

拉普拉斯导出他所谓的“接续法则”，该法则用于估算太阳升起n次后会再次升起的概率，表示为  $(n+1)/(n+2)$ 。当n=0时，这个概率为1/2；随着n增加，概率也会增加，当n接近无穷大时，概率接近1。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2460-2463. Kindle Edition

Notes: 1) 接续法则

我们将其称为太阳会升起来的“先验概率”，因为这发生在看到任何证据之前。“先验概率”的基础并不是数过去这个星球上太阳升起的次数，因为过去你没有看到；它反映的是对于将要发生的事情，你优先相信的东西，这建立在你掌握的宇宙常识之上。但现在星星开始渐渐暗淡，所以你对于太阳会升起的信心越来越强，这建立于你在地球上生存的经历之上。你的这种信心源自“后验概率”，因为这个概率是在看到一些证据后得出的。天空开始渐渐变亮，后验概率又变得更大了。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2466-2470. Kindle Edition

$P(\text{原因} | \text{结果}) = P(\text{原因}) \times P(\text{结果} | \text{原因}) / P(\text{结果})$  用A代替原因，用B代替结果，然后为了简洁，把乘法符号删掉，你就会得到大教堂上那个用10英尺宽的字母书写的公式。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2485-2488. Kindle Edition

这就是贝叶斯定理，感冒是原因，发烧是结果。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2501-2502. Kindle Edition

贝叶斯定理让我们由原因推出结果，又由结果知道原因，

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2512-2513. Kindle Edition

对于贝叶斯学派来说，没有所谓的真相。你有一个优先于假设的分布，在见到数据后，它变成了后验分布，这是贝叶斯定理给出的说法，也就是贝叶斯定理的全部。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2812-2814. Kindle Edition

叶斯学派指出了逻辑的脆弱性。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2914-2915. Kindle Edition

贝叶斯学派指出了逻辑的脆弱性。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2914-2915. Kindle Edition

我们既需要逻辑，也需要概率。治愈癌症就是一个很好的例子。贝叶斯网络可以从单个方面模仿细胞如何起作用，就像基因调节和蛋白质折叠那样，但只有逻辑可以将所有碎片组合到一张连贯的图片中。此外，逻辑无法处理不完整或包含嘈杂因素的信息，这在实验生物学中较普遍，但贝叶斯网络可以沉着地处理这个问题。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2927-2930. Kindle Edition

但肯尼迪那时刚读过《八月炮火》，这是一本描写第一次世界大战的畅销书，所以他知道那样做很容易会引发全面战争。所以他选择了对古巴进行海上封锁，也许这样做把世界从核战争中拯救出来了。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 2962-2965. Kindle Edition

1994年，明尼苏达州立大学和麻省理工学院的研究人员建立了一个推荐系统，其构建基础是他们所谓的“一个看似简单的想法”：过去人们同意的话，将来他们也还会同意。这个想法直接引出协同过滤系统，所有典型的电子商务网站都有这些系统。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3044-3046. Kindle Edition

一些研究人员甚至认为，创造智能机器的方法，就是要造出一个机器婴儿，让它像人类婴儿那样去体验世界，研究人员就是它的父母（甚至可能还会有来自众包的协助，赋予“地球村”这个术语以全新的含义）。小

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3341-3343. Kindle Edition

一些研究人员甚至认为，创造智能机器的方法，就是要造出一个机器婴儿，让它像人类婴儿那样去体验世界，研究人员就是它的父母（甚至可能还会有来自众包的协助，赋予“地球村”这个术语以全新的含义）。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3341-3343. Kindle Edition

你可能已经注意到k均值算法和EM之间的某个相似性，因为它们都交替进行两项工作：将实体分配给集群，然后更新集群的描述。这并不是一场意外：k均值本身就是EM的一种特殊情况，当所有属性都会有“狭窄的”正态分布时（有很小变量的正态分布），你就会找到它。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3434-3437. Kindle Edition

理解并控制由线索、日常、奖励组成的循环关系是成功的关键，不仅对个人，而且对企业甚至整个社会来说都是这样。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3618-3619. Kindle Edition

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3633. Kindle Edition

这种类型的曲线被称为“幂法则”（a power law），因为随着时间达到某负幂时，表现会出现变化。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3633-3634. Kindle Edition

人类解决问题的方式是将问题分解为小问题，再将小问题再分解为更小的问题，然后系统地减少初始状态（比如第一个公式）与目标状态（第二个公式）之间的差异。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3649-3650. Kindle Edition

H·L·门肯说过一句妙语：如果一个人挣的钱比他的妻子的妹妹的丈夫要多，那么他就是富有的。这句话涉及四个人。这些规则中的每一条都可以变成关联模型中的特征模板，而在特征出现在数据中的频率的基础上，可以确定它的一个权值。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3722-3724. Kindle Edition

口碑传播。信息如何在社交网络中传播？我们可否测量每个成员的影响力，然后将目标确定在数量够多、影响力最大的成员身上，以开启一轮口头传播？

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3745-3746. Kindle Edition

在艾萨克·阿西莫夫的《基地》一书中，科学家哈里·谢顿可以从数学角度预测人类的未来，从而将其从衰落中拯救出来。此外，保罗·克鲁格曼也承认，就是这个诱人的梦想，让他成为一名经济学家。依据谢顿的观点，人类就像气体中的分子，大数法则保证，即使个人无法预测，整个社会却可以。关联学习表明事实并非如此，如果人类是独立的，每个人孤立地做决定，社会的确就会变得可预测，因为所有那些随意的决定会合计为一个相当恒定的均值。但当人们互动时，较大的集合体会比较小的集合体更不那么可预测。如果信心和恐惧可以传染，每种情绪都会主导一段时间，但常常整个社会会从这种情绪转化到另一种情绪。虽然如此，并不是所有消息都是坏的。如果我们估计人们之间相互影响的强度，也就可以估计这种转换在多久之后会发生，即使这是第一次转换。换个说法，黑天鹅并不一定不可预测。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3751-3759. Kindle Edition

在科学方面，理论的统一往往从看似简单的观察开始。两个看似不相关的现象原来只是同一枚硬币的两面，就像第一张倒下的多米诺骨牌，会引起其他许多牌倒下。苹果落到地上，月亮悬挂在夜空，这两者都是由重力引起的，而且（无论是否杜撰）一旦牛顿弄明白这些现象的原因，重力最后也可以用来解释潮汐、分点岁差、彗星轨道等其他很多现象。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3778-3781. Kindle Edition

甚至电也是一种统一物：你可以通过许多不同的来源来获取它——煤、天然气、核能、水力、风力、太阳，然后以无限多种的方式来消耗它。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3799-3801. Kindle Edition

甚至电也是一种统一物：你可以通过许多不同的来源来获取它——煤、天然气、核能、水力、风力、太阳，然后以无限多种的方式来消耗它。一座发电站不会知道或者关心它生产的电会如何被消耗掉，而你的门廊灯、洗碗机或者全新的特斯拉也不会在意电力供应来自哪里。电力就是能源的世界语言。终极算法是机器学习的统一物：它让任意应用利用任意学习算法，方法是将学习算法概括成通用形式——所有应用都需要知道该形式。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3799-3804. Kindle Edition

网飞奖获得者利用元学习来结合数百个不同的学习算法；沃森利用它来从备选项中选择最终的答案；内特·希尔也以相似的方式将投票与预测选举结果结合起来。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3818-3820. Kindle Edition

最聪明的元学习算法之一就是推进，由两位学习领域的理论家约阿夫·弗罗因德和罗伯·夏皮尔创造。

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3827-3828. Kindle Edition

佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3847. Kindle Edition

你专注地盯着地图，  
佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3848-3848. Kindle Edition

符号学派的形式语言是逻辑，其中规则和决策树是特殊例子。联结学派的是神经网络。进化学派的是遗传程序，包括分类器系统。贝叶斯学派的是图解模型，这是贝叶斯网络和马尔可夫网络的涵盖性术语。类推学派的是特例，可能会有权值，就像在支持向量机中那样。  
佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3851-3854. Kindle Edition

优化是一种算法，即寻找最高得分的模型，并回归它。符号学派的特色搜索算法是逆向演绎。联结学派的是梯度下降。进化学派的是遗传搜索，包括交叉和突变。贝叶斯学派在这方面不同寻常：他们不只是寻找最好的模型，而是寻找所有模型的平均值，由它们的可能程度来权衡。为了有效进行加权，他们利用诸如MCMC之类的推理算法。类推学派（或者更准确地说是支持向量机）利用约束优化来找到最佳模型。  
佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3857-3861. Kindle Edition

有5个大门，每个门由一个学派掌控，并通往  
佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3862-3863. Kindle Edition

思考如何将逻辑和概率统一起来，  
佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3941-3941. Kindle Edition

我们到达的统一学习算法利用MLN作为表示方法，利用后验概率作为评估函数，利用与梯度下降结合的基因搜索作为优化器。如果我们愿意，可以轻易地利用其他准确度测量方法来代替后验概率，或者利用爬山法来代替遗传搜索。我们上升到一座高峰，现在我们可以享受风景了。  
佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 3991-3993. Kindle Edition

我们可以掌握世界的MLN，方法就是将其分解成若干部分和子部分（这样多数互动关系就存在于同一部分的子部分之间），然后将各个部分集成成类别和子类别。如果世界是一个乐高玩具，那么我们将其分解为一块块“砖”，记住哪块和哪块相连，然后通过形状和颜色来将“砖块”集合起来。如果这个世界是维基百科，我们可以将它谈论的实体抽取出来，将它们集成成类别，然后学习类别之间如何相互关联。  
佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 4114-4118. Kindle Edition

花点时间来考虑你记录在世界上所有计算机里的数据：你的邮箱、办公文档、文本；推特、脸书和领英账号；你的网页搜索、点击、下载、购买；你的信用卡、传真、电话、健康档案；你的健康追踪器统计；你的汽车微处理器记录下的驾驶情况；你闲逛时被手机记录下来的信息；你拍过的所有照片；监控摄像机里的简短片段；你的谷歌眼镜片段。  
佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 4247-4250. Kindle Edition

该算法已经包含所有我们教过它的所有东西。  
佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 4256-4257. Kindle Edition

注意到，其中会有一个收集你的数据的疯狂比赛。每个人都喜欢你的数据，这也难怪，它们是通往你的世界、你的钱包、你的投票甚至你的心灵的大门。但是每个人只能拥有它的一小部分：谷歌掌握你搜索的内容，亚马逊知道你网购的东西，美国电话电报公司会看到你的通话记录，苹果知道你下载的音乐，西夫韦懂得你购买的杂货，美国第一资本投资国际集团了解你的信用卡交易记录。诸如安客诚（Acxiom）之类的公司会整理并销售关于你的数据，但如果你可以对其进行检查（对于安客诚的情况，你可以在aboutthedata.com检查），数据并不多，而且有些还是错误的。没有人能够了解到完完整整的你。这有好处，也有坏处。有好处是因为如果某人做到了，他就会掌握很大的权力；有坏处是因为只要事实是那样的，就不会有你的360°模型了。  
佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 4320-4327. Kindle Edition

公司和个人相比，收集和使用数据的能力要强大很多，这导致了权利上的不平衡。数据越有价值，就越能从中掌握更好、更有用的模型，不对称也就越严重。  
佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 4365-4366. Kindle Edition

你的工作会在多大程度上用到你的大脑？用得越多，你就越安全。  
佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 4385-4385. Kindle Edition

如果计算机已经学会完成你的工作，不要试图与它竞争，而要利用它。  
佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 4398-4398. Kindle Edition

随着技术的进步，人和机器更加密切的结合体就形成了：你饿了，Yelp会推荐一些好吃的餐厅；GPS会指引你方向；你开车，汽车电子会进行低水平控制。我们现在都已经是半机器人了。真正的自动化指的不是它代替了什么，而是它增强了什么能力。  
佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 4404-4406. Kindle Edition

早前列举的机器学习对日常生活的影响可在George John（SIGKDD Explorations, 1999）所著的“Behind-the-scenes data mining”中找到，序言部分“日常生活”段落的灵感也源于此。  
佩德罗·多明戈斯, 终极算法:机器学习和人工智能如何重塑世界, loc. 4672-4673. Kindle Edition

Notes: 1) 本书引用的资源