

局部连接来减参，权值共享肩并肩（深度学习入门系列之十一）(<https://yq.aliyun.com/articles/159710>)

激活引入非线性，池化预防过拟合（深度学习入门系列之十二）(<https://yq.aliyun.com/articles/167391>)

循环递归RNN，序列建模套路深（深度学习入门系列之十三）(<https://yq.aliyun.com/articles/169880>)

LSTM长短记，长序依赖可追忆（深度学习入门系列之十四）(<https://yq.aliyun.com/articles/174256>)

云栖社区微信小程序“云栖365”上线了，帮助碎片时间学习 (/roundtable/495548)

你的教程你做主！想要什么样的大数据开发？ (/roundtable/495338)

三七女生节，邀你一起解密阿里女程序员们 (/roundtable/495243)

阿里巴巴航母级计算平台MaxCompute即将‘者’版本，与开发者共建大数据生态 (/roundtable/495195)

数据千万条，备份第一条，数据找不回，老 (/roundtable/494595)

【导言】目前人工智能非常火爆，而深度学习则是引领这一火爆现场的“火箭”。于是，有关“深度学习”的论文、书籍和网络博客汗牛充栋，但大多数这类文章都具备“高不成低不就”的特征。对于高手来说，自然是没有问题，他们本身已经具备非常“深度”的学习能力，如果他们想学习有关深度学习的技术，直接找来最新的研究论文阅读就好了。但是，对于低手（初学者）而言，就没有那么容易了，因为他们基础相对薄弱，通常看不太懂。

于是，我们推出深度学习的入门系列。在这个系列文章中，我们力图用最通俗易懂、图文并茂的方式，带你入门深度学习。我们都知道，**高手从来都是自学出来的**。所以，这个深度学习的入门系列，能带给你的是“从入门到精通”，还是“从入门到放弃”，一切都取决于你个人的认知。**成就你自己的，永远都是你自己**，是吧？

好了，言归正传，下面开始我们的正题。

## 1.1 什么是学习？

说到深度学习，我们首先需要知道，什么是学习。

著名学者赫伯特·西蒙教授（Herbert Simon，1975年图灵奖获得者、1978年诺贝尔经济学奖获得者）曾对“学习”给了一个定义：“如果一个系统，能够通过执行某个过程，就此改进了它的性能，那么这个过程就是学习”。

大牛就是大牛，永远都是那么言简意赅，一针见血。从西蒙教授的观点可以看出，**学习的核心目的，就是改善性能**。

其实对于人而言，这个定义也是适用的。比如，我们现在正在学习“深度学习”的知识，其本质的就是为了“提升”自己在机器学习上的认知水平。如果我们仅仅是低层次的重复性学习，而没有达到认知升级的目的，那么即使表面看起来非常勤奋，其实我们也仅仅是个“伪学习者”，因为我们没有改善性能。

## 1.2 什么是机器学习？

遵循西蒙教授的观点，对于计算机系统而言，通过运用数据及某种特定的方法（比如统计的方法或推理的方法），来提升机器系统的性能，就是机器学习。

英雄所见略同。卡内基梅隆大学的Tom Mitchell教授，在他的名作《机器学习》一书中，也给出了更为具体（其实也很抽象）的定义[1]：

对于某类任务（Task，简称T）和某项性能评价准则（Performance，简称P），如果一个计算机程序在T上，以P作为性能的度量，随着很多经验（Experience，简称E）不断自我完善，那么我们称这个计算机程序在从经验E中学习了。

比如说，对于学习围棋的程序AlphaGo，它可以通过和自己下棋获取经验，那么它的任务T就是“参与围棋对弈”；它的性能P就是用“赢得比赛的百分比”来度量。“类似地，学生的任务T就是“上课看书写作业”；它的性能P就是用“期末成绩”来度量”



(<https://www.aliyun.com/acts/product-section-2019/developer>)

因此，Mitchell教授认为，对于一个学习问题，我们需要明确三个特征：任务的类型，衡量任务性能提升的标准以及获取经验的来源。

### 1.3 学习的4个象限

在前面的文章中[2]，我们已提到，一般说来，人类的知识在两个维度上可分成四类（见图1-1）。即从可统计与否上来看，可分为：是可统计的和不可统计的。从能否推理上看，可分为可推理的和不可推理的。

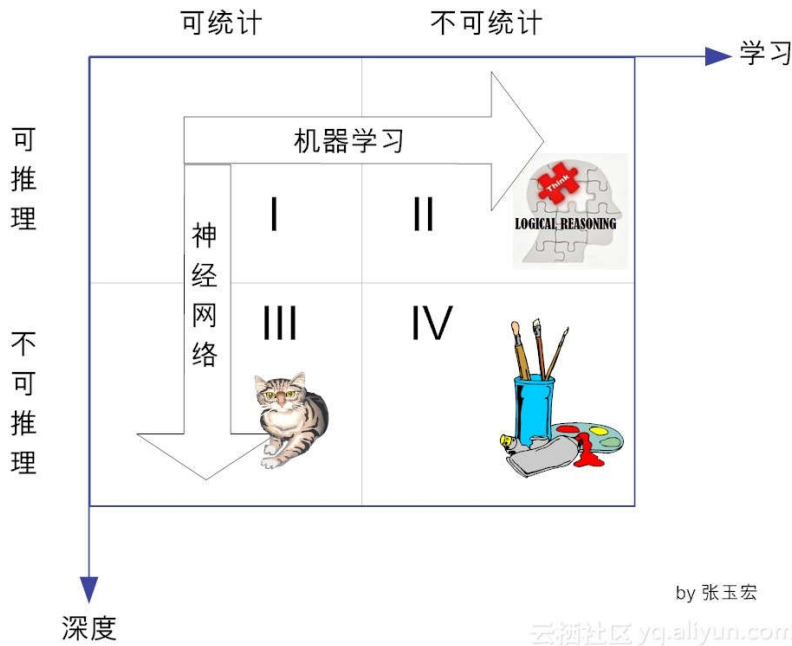


图1-1 人类知识的4个象限

在横向方向上，对于可推理的，我们都可以通过机器学习的方法，最终可以完成这个推理。传统的机器学习方法，就是试图找到可举一反三的方法，向可推理但不可统计的象限进发（象限II）。目前看来，这个象限的研究工作（即基于推理的机器学习）陷入了不温不火的境地，能不能峰回路转，还有待时间的检验。

而在纵向上，对于可统计的、但不可推理的（即象限III），可通过神经网络这种特定的机器学习方法，以期望达到性能提升的目的。目前，基于深度学习的棋类博弈（阿尔法狗）、计算机视觉（猫狗识别）、自动驾驶等等，其实都是在这个象限做出了了不起的成就。

从图1可知，深度学习属于统计学习的范畴。用李航博士的话来说，统计机器学习的对象，其实就是数据[3]。这是因为，对于计算机系统而言，所有的“经验”都是以数据的形式存在的。作为学习的对象，数据的类型是多样的，可以是各种数字、文字、图像、音频、视频，也可以是它们的各种组合。

统计机器学习，就是从数据出发，提取数据的特征（由谁来提取，是个大是大非问题，下面将给予介绍），抽象出数据的模型，发现数据中的知识，最后又回到数据的分析与预测当中去。

### 1.4 机器学习的方法论

这里稍早说明的一点的是，在深度学习中，经常有“end-to-end（端到端）”学习的提法，与之相对应的传统机器学习是“Divide and Conquer（分而治之）”。这些都是什么意思呢？

“end-to-end”（端到端）说的是，输入的是原始数据（始端），然后输出的直接就是最终目标（末端），中间过程不可知，因此也难以知。比如说，基于深度学习的图像识别系统，输入端是图片的像素数据，而输出端直接就是或猫或狗的判定。这个端到端就是：像素→判定。

再比如说，“end-to-end”的自动驾驶系统[4]，输入的是前置摄像头的视频信号（其实也就是像素），而输出的直接就是控制车辆行驶指令（方向盘的旋转角度）。这个端到端就是：像素→指令。

就此，有人批评深度学习就是一个黑箱（Black Box）系统，其性能很好，却不知道为何而好，也就是说，缺乏解释性。其实，这是由于深度学习所处的知识象限决定的。从图1可以看出，深度学习，在本质上，属于可统计不可推理的范畴。“可统计”是很容易理解的，就是说，对于同类数据，它具有一定的统计规律，这是一切统计学习的基本假设。那“不可推理”又是什么概念？其实就是“剪不断、理还乱”的非线性状态了。



图1-2 1个人+个人1=?

在哲学上讲，这种非线性状态，是具备了整体性的“复杂系统”，属于复杂性科学范畴。复杂性科学认为，构成复杂系统的各个要素，自成体系，但阡陌纵横，其内部结构难以分割。简单来说，对于复杂系统， $1+1 \neq 2$ ，也就是说，一个简单系统，加上另外一个简单系统，其效果绝不是两个系统的简单累加效应，而可能是大于部分之和。因此，我们必须从整体上认识这样的复杂系统。于是，在认知上，就有了从一个系统或状态（end）直接整体变迁到另外一个系统或状态（end）的形态。这就是深度学习背后的方法论。

与之对应的是“Divide and Conquer（分而治之）”，其理念正好相反，在哲学它属于“还原主义（reductionism，或称还原论）”。在这种方法论中，有一种“追本溯源”的蕴意包含其内，即一个系统（或理论）无论多复杂，都可以分解、分解、再分解，直到能够还原到逻辑原点。

在意象上，还原主义就是“ $1+1=2$ ”，也就是说，一个复杂的系统，都可以由简单的系统简单叠加而成（可以理解为线性系统），如果各个简单系统的问题解决了，那么整体的问题也就得以解决。比如说，很多的经典力学问题，不论形式有多复杂，通过不断的分解和还原，最后都可以通过牛顿的三大定律得以解决。

经典机器学习（位于第Ⅱ象限），在哲学上，在某种程度上，就可归属于还原主义。传统的机器学习方式，通常是用人类的先验知识，把原始数据预处理成各种特征（feature），然后对特征进行分类。

然而，这种分类的效果，高度取决于特征选取的好坏。传统的机器学习专家们，把大部分时间都花在如何寻找更加合适的特征上。因此，早期的机器学习专家们非常苦逼，故此，传统的机器学习，其实可以有个更合适的称呼——特征工程（feature engineering）。

但这种苦逼，也是有好处的。这是因为，这些特征是由人找出来的，自然也就为人所能理解，性能好坏，机器学习专家们可以“冷暖自知”，灵活调整。

### 1.5 什么是深度学习？

再后来，机器学习的专家们发现，可以让神经网络自己学习如何抓取数据的特征，这种学习的方式，效果更佳。于是兴起了特征表示学习（feature representation learning）的风潮。这种学习方式，对数据的拟合也更加的灵活好用。于是，人们终于从自寻“特征”的苦逼生活中解脱出来。

但这种解脱也付出了代价，那就是机器自己学习出来的特征，它们存在于机器空间，完全超越了人类理解的范畴，对人而言，这就是一个黑盒世界。为了让神经网络的学习性能，表现得更好一些，人们只能依据经验，不断地尝试性地进行大量重复的网络参数调整，同样是“苦不堪言”。于是，“人工智能”领域就有这样的调侃：“有多少人工，就有多少智能”。

因此，你可以看到，在这个世界上，存在着一个“麻烦守恒定律”：**麻烦不会减少，只会转移。**

再后来，网络进一步加深，出现了多层次的“表示学习”，它把学习的性能提升到另一个高度。这种学习的层次多了，其实也就是套路“深了”。于是，人们就给它取了个特别的名称——Deep Learning（深度学习）。

深度学习的学习对象同样是数据。与传统机器学习所不同的是，它需要大量的数据，也就是“大数据（Big Data）”。

有一个观点，在工业界一度很流行，那就是在大数据条件下，简单的学习模型会比复杂模型更加有效。而简单的模型，最后会趋向于无模型，也就是无理论。例如，早在2008年，美国《连线》（Wired）杂志主编克里斯·安德森（Chris Anderson）就曾发出“理论的终结（The End of Theory）”的惊人断言[5]：“海量数据已经让科学方法成为过去时（The Data Deluge Makes the Scientific Method Obsolete）”。

但地平线机器人创始人（前百度深度学习研究院副院长）余凯先生认为[6]，深度学习的惊人进展，是时候促使我们要重新思考这个观点了。也就是说，他认为“大数据+复杂模型”或许能更好地提升学习系统的性能。

### 1.6 “恋爱”中的深度学习

法国科技哲学家伯纳德·斯蒂格勒（Bernard Stiegler）认为，人们以自己的技术和各种物化的工具，作为自己“额外”的器官，不断的成就自己。按照这个观点，其实，在很多场景下，计算机都是人类思维的一种物化形式。换句话说，计算机的思维（比如说各种电子算法），都能找到人类生活实践的影子。

比如说，现在火热的深度学习，与人们的恋爱过程也有相通之处。在知乎上，就有人（jacky yang）以恋爱为例来说明深度学习的思想，倒也非常传神。我们知道，男女恋爱大致可分为三个阶段：

第一阶段初恋期，相当于深度学习的输入层。妹子吸引你，肯定是有许多因素，比如说脸蛋、身高、身材、性格、学历等等，这些都是输入层的参数。对不同喜好的人，他们对输出结果的期望是不同的，自然他们对这些参数设置的权重也是不一样的。比如，有些人是奔着结婚去的，那么他们对妹子的性格可能给予更高的权重。否则，脸蛋的权重可能会更高。

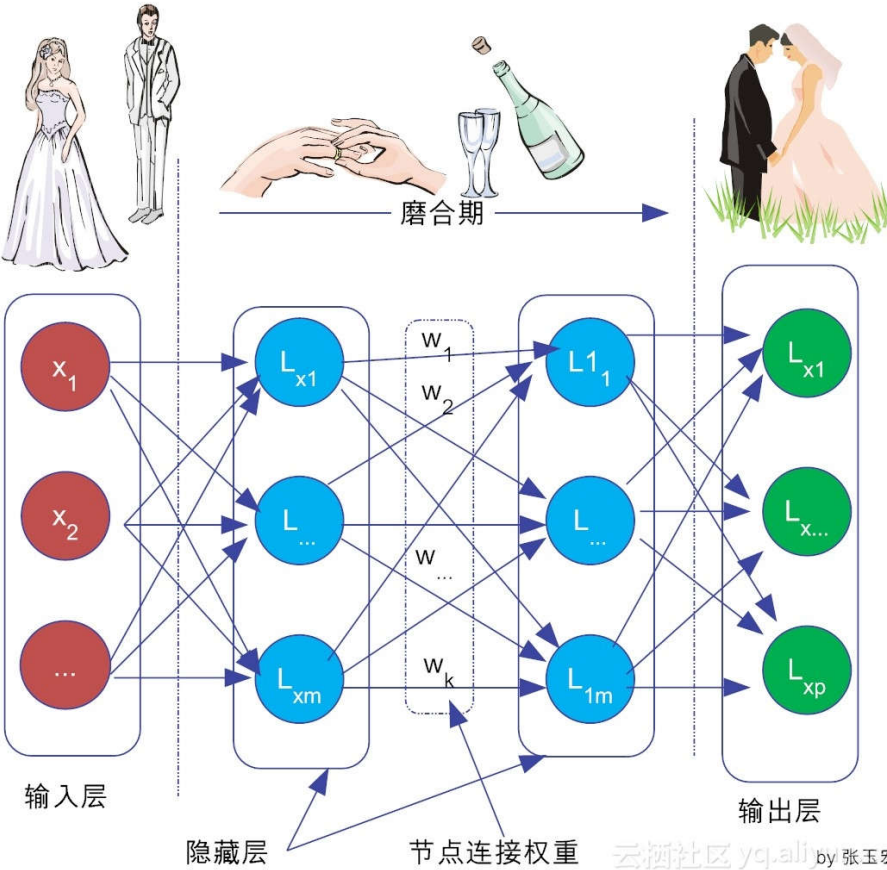


图1-3 恋爱中的深度学习

第二阶段热恋期，对应于深度学习的隐藏层。在这个期间，恋爱双方都要经历各种历练和磨合。清朝湖南湘潭人张灿写了一首七绝：

书画琴棋诗酒花，当年件件不离他。  
而今七事都变更，柴米油盐酱醋茶。

这首诗说的就是，在过日子的洗礼中，各种生活琐事的变迁。恋爱是过日子的一部分，其实也是如此，也需要双方不断磨合。这种磨合中的权重取舍平衡，就相等于深度学习中隐藏层的参数调整，它们需要不断地训练和修正！恋爱双方相处，磨合是非常重要的。要怎么磨合呢？光说“520（我爱你）”，是廉价的。这就给我们程序猿（媛）提个醒，爱她（他），就要多陪陪她（他）。陪陪她（他），就增加了参数调整的机会。参数调整得好，输出的结果才能是你想要的。

第三阶段稳定期，自然相当于深度学习的输出层。输出结果是否合适，是否达到预期，高度取决于“隐藏层”的参数“磨合”得怎么样。

1.7 小结

在本小节，我们回顾了“机器学习”的核心要素，那就是通过对数据运用，依据统计或推理的方法，让计算机系统的性能得到提升。而深度学习，则是把由人工选取对象特征，变更为通过神经网络自己选取特征，为了提升学习的性能，神经网络的表示学习的层次较多（较深）。

以上仅仅给出机器学习和深度学习的概念性描述，在下一个小结中，我们将给出机器学习的形式化表示，传统机器学习和深度学习的不同之处在哪里，以及到底什么是神经网络等。

1.8 请你思考

- (1) 在大数据时代，你是赞同科技编辑出生的克里斯·安德森的观点呢（仅需简单模型甚至无模型），还是更认可工业界大神余凯先生的观点呢（还是需要复杂模型）？为什么？
- (2) 你认为用“恋爱”的例子比拟“深度学习”贴切吗？为什么？
- (3) 为什么非要用“深度”学习，“浅度”不行吗？