人机对话续

万精油

编者按：十二年前，本刊刊载了《人机对话》[1] 一文，介绍了当时的人工智能，尤其是在围棋程序方面的人工智能的发展及前景，反映了人们在这条道路上的一些探索与努力。从了解人的思维以及先驱们的探索角度，那篇文章至今仍有一定价值。但是，在现在一切都以摩尔定律发展的年代，十几年前的东西注定是会有许多过时的，因此，原作者将这十几年间的一些情况补充了进来，也就是这篇《人机对话续》，使读者能够对这方面的发展有更新的认识。

AI 围棋的分水岭是阿尔法狗（AlphaGo）。《人机对话》讲的是阿尔法狗以前的情况，这篇续就来讲讲阿尔法狗以及它后面的情况。

我们先简述一下阿尔法狗以前的情况。

人工智能的研究从上世纪五十年代就开始了。这个研究的一个试金石就是智能对弈，机器挑战人类。最早是简单的五子棋，跳棋，之类的。到 1997 年 IBM 的深蓝战胜了国际象棋世界冠军，算是一个很大的里程碑。但是，当时

[1] 数学文化 , 2011/ 第 2 卷第 1 期 , pp 85-90.

的算法、硬件、软件都还不能撼动围棋。不要说世界冠军，就连一般的业余围棋棋手都可以让它好几子。这里面关键的问题是，围棋选择点太多，而且，没有一个很好的鉴定函数来判断什么是好棋，什么是坏棋。有人想到用蒙特卡罗方法来对付这个问题，没法判定好坏就一直走到底，总可以判定胜负。把所有招数都走到底是不可能的，蒙特卡罗方法是随机模拟，随机地选一些位置来走，如果某个点走下去赢的比例最大，那么就选它为下一步。这个方法出乎意料地有用。用这个方法搞出来的围棋程序已经有业余高手的水平，甚至可以在9×9 棋盘上胜职业棋手。

但是正规 19×19 的棋盘上选点太多，没有选择性地模拟出来的点与职业棋手下出来的棋相去甚远，更没有可能战胜世界冠军。

原来的共识是，按照当时的进度，战胜人类世界冠军的围棋程序至少要在几十年后才会出现（有悲观主义者甚至认为这样的程序永远不会出现）。但是，2016 年 1 月，阿尔法狗横空出世，改变了一切。

事件回放：2015 年 10 月，谷歌深思集团领军人物哈萨比斯（Demis Hassabis）在接受采访时说，几个月内会有重大消息宣布。2016 年 1 月，谷歌深思集团在《自然》杂志发文，介绍了阿尔法狗（Alpha Go）的原理和方法，并刊登了阿尔法狗 5:0 战胜职业棋手樊麾的五盘棋谱，并表示它已经能挑战人类顶尖棋手，韩国的围棋世界冠军李世石已经接受了挑战。

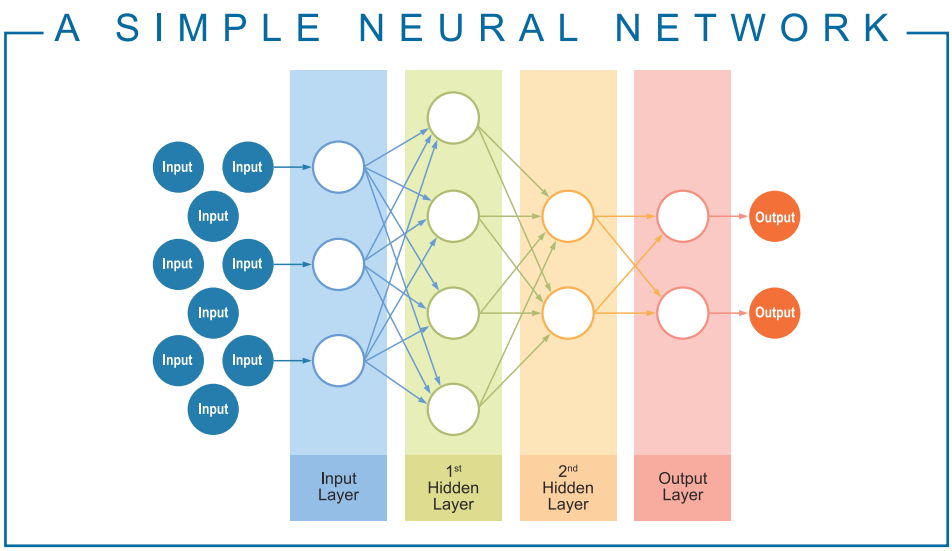
同年三月份的阿尔法狗与李世石的五番棋可谓万众瞩目，几乎所有的职业棋手都不看好阿尔法狗，都认为李世石会赢。但是，结果让他们很失望，李世石以 1:4 落败。围棋和国际象棋曾经被认为是人类能抵抗计算机的最后两个堡垒。1997 年 IBM 的深蓝击败国际象棋大师卡斯帕洛夫以后，围棋成了唯一的堡垒。当时有记者问深蓝的领军人物许峰雄，是否要开始攻克围棋，他的回答是，围棋太难了，这个堡垒还会存在一段时间。李世石的落败宣告了这最后一块堡垒的垮塌。

很自然的一个问题是，阿尔法狗是靠什么攻克围棋的？它与之前的围棋程序有什么不同？

如果用一个词来总结阿尔法狗的精髓的话，那就是“深度学习”。通过神经网络，学习专家棋谱。

神经网络作为生物概念在 19 世纪就有了。我们人类的大脑就是神经元加上一些连接。大脑工作的时候就是神经元通过这些连接传递信息（信号），处理，再传递，最后得出指令。比如一个场景通过视网膜传进大脑发向许许多多的神经元，再传递给与其连接的神经元，处理分解后再传下去。如果看见楼梯，最后的指令就是抬腿。如果看见羽毛球飞过来，指令就是挥拍。人工智能所指的神经网络概念要到上世纪 50 年代才出现，其宗旨就是用很多节点模拟大脑神经元，用这些节点之间的连接模拟神经元之间的连接。所谓“机器学习”就是不断更新这些连接的权重。最后收敛成一个能输出好结果的黑匣子。信息从一端输入，通过这些连接传递到最后得出一个指令。对于下围棋来说，输入的信息就是当前的棋盘，输出的指令就是得出下一步该下哪里。人工智能的神经网

络刚开始出现的时候，中间的节点只有一两层。随着计算机的逐渐发达，层数越来越多，“深度学习”里的深度就是层数很多的意思。阿尔法狗用了 13 层。



在阿尔法狗出现以前，神经网络比较成功的例子是图像识别。识别手写字，识别图像里的猫、狗等等。阿尔法狗用的就是图像识别的手法。

以前的围棋程序，一个棋盘用一个 19×19 的矩阵表示，矩阵的每个元素又有很多属性，比如是否有子，有的话是什么颜色，有多少口气，是否处于打劫状态等等。阿尔法狗不用这些，直接把棋谱当成一个 19×19 个像素的图像。输入目前的图像，通过学习，找出下一个图像（下一步棋）。

神经网络是如何学习的呢？神经网络的结构建好以后，输入一个图像，通过中间层的各种映射最后到终端可以输出一个结果（比如下一步棋的图像，或者现有棋盘上空格点作为下一步的概率）。开始的时候，中间层的链接（映射）是随机的，输出的结果也没有章法。但我们可以检查这个结果与标准答案或者最佳结果的差距来优化中间层的映射，逐渐向最小差距收敛。这就是一个学习过程。

这里面有两个问题。第一个问题是，什么是标准答案？不同的学习方法有不同的标准答案。阿尔法狗采用的是从围棋网站上下载的三千万个围棋高段（大多是职业棋手）的棋谱生成的图片。这种有专家知识辅助的学习方法叫监督学习（supervised learning）。还有一种没有专家知识的学习方法叫强化学习（reinforcement learning），对阿尔法狗来说，就是自我对弈，然后从结果判断哪条路更优秀。

第二个问题是，如何优化。因为有些中间层中有非线性的映射。我们可以把一般数值分析中使用的优化手法用到这里。求导数，然后沿梯度方向前进，把差距极小化。

通过学习，阿尔法狗产生了两个神经网络。一个叫策略网络（Policy