机器下棋的历史与启示
——从“深蓝”到AlphaZero

薛永红[1,2],王洪鹏[3]

1. 华北科技学院理学院,北京 101601
2. 北京师范大学哲学学院,北京 100875
3. 中国科学技术馆,北京 100012

摘要　以历史为线索,从设计思路和技术特征两个方面对“深蓝”和AlphaGo进行了梳理和概括。“深蓝”依赖人类在国际象棋领域的经验,借助强大的算力与算法实现了对人类的超越;20年后的AlphaGo,虽然最初的版本也是利用人类经验而获得成功的,但是它的不断进化却揭示了一个重要事实:人类经验具有局限性。放弃人类经验、完全采用机器自对弈经验的AlphaZero,不但具有最强的围棋对弈能力,而且同时具备国际象棋和日本将棋的最高棋力,3种最强技能集于一身。机器下棋的这一历史线索揭示了在棋类游戏中,囿于人类自身认知能力的局限,人类几千年积累下来的经验较之于机器在短期内所形成的“经验”已不占优势。在巨大的算力和不断完善的算法的支撑下,借助于机器自身“经验”,机器可以做得比人类更好。未来,“放弃人类经验,依靠自身经验”的机器将有可能在更为复杂的领域取得突破性进展。

关键词　机器下棋;深蓝;AlphaGo;AlphaZero

棋类游戏一直是人工智能所要攻克的领域。1947年,图灵(A. L. Turing)编写了一个国际象棋的程序,但是由于计算机在当时是稀缺资源,使得这个程序没有机会在计算机上运行。与此同时,信息理论的创始人香农(C. E. Shannon)等提出了双人对弈的最小最大算法(Minimax),并于1950年发表了理论研究论文《Programming a computer for playing chess》(《计算机下棋程序》)[1],首开理论研究机器下棋的先河。在文中,棋盘被定义为一个二维数组,每个棋子都被赋予一个子程序,用于对棋子可能走法的计算,当子程序计算出所有可能的走法后,就会得到一个评估函数,用每个棋子的可能走法就可以形成一个博弈树。对于一个完全信息的博弈系统,如果能穷举完整的博弈树,那Mini-

收稿日期:2019-07-05;修回日期:2019-08-23

基金项目:2018年度国家社会科学基金重点项目(18AZX008);中央高校基本科研业务费项目(3142018057)

作者简介:薛永红,副教授,研究方向为科学思想史与科学社会史,电子信箱:aristotle@ncist.edu.cn

引用格式:薛永红, 王洪鹏. 机器下棋的历史与启示——从“深蓝”到AlphaZero[J]. 科技导报, 2019, 37(19): 87-96; doi: 10.3981/j.issn.1000-7857.2019.19.012

max算法就可以计算出最优的策略[2]。由于复杂游戏的博弈树增长是指数形式的,因此要穷举完整的博弈树非常困难。约翰·麦卡锡(J. McCarth)提出了著名的α-β剪枝技术,对有效控制博弈树的规模提供了依据。随后,卡内基梅隆大学的纽厄尔(A. Newell)、司马贺(H. Simon)等很快在实战中实现了这一技术。Minimax算法必须在完成完整的博弈树之后才能计算评估函数,而α-β剪枝技术则是一边画博弈树,一边进行计算,一旦在计算过程中评估函数出现"溢出",则自动停止对树的进一步搜索,从而极大地减小了博弈树的规模和实际的搜索空间。这一创新被看作是攻克棋类游戏的重要法宝,并且首先在国际象棋领域大获成功。

1 "深蓝"成功的秘诀

起初,由于机器下棋的水平远达不到人类普通棋手,所以比赛一般都是在机器之间进行。1989年,卡内基梅隆大学的团队开发出的下棋机"深思"(Deep Thought),成为第1个国际象棋的计算机特级大师。此后这个团队加入IBM,成为后来"深蓝"(Deep Blue)的核心团队。1997年5月,在美国纽约举行的一场六局的比赛中,"深蓝"战胜了卡斯帕罗夫,从而成为历史上第1个战胜人类国际象棋大师的下棋机。

与卡斯帕罗夫对战的"深蓝"有2个操作台,包括30台计算机(路机),其中用到了480个定制的国际象棋芯片[3]。因此,可以将"深蓝"视为通过高速交换网络连接的IBM RS/6000处理器或工作站的集合。而IBM开发的用于RS/6000的超能2芯片(P2SC)可以使SP2计算机以130 MHz的速度运行。"深蓝"系统中的每个处理器最多可控制16个国际象棋芯片,分布在2个微信道卡上,每张卡上有8个国际象棋芯片,有超过4000个的处理器[4]。这一系统每秒可以检索2×10^8个棋局,而且检索的深度也有了进一步提高。因此,专用芯片所提供的强大算力为"暴力穷举"算法的实现提供了基础。加上丰富的象棋知识、残局、改进的开局库以及在特级大师的仔细检验下进行了1年的测试,最终版本的"深蓝"棋力非常之强。

可以看出,"深蓝"能战胜国际象棋大师,主要是基于两点:第一是丰富的国际象棋知识,尤其是对这些知识的深入理解[5];第二是巨大的算力。虽然剪枝算法以及软件对残局的搜索客观上降低了搜索空间,但整体上依然属于暴力穷举。这种设计思路即使是在20年后的围棋下棋机AlphaGo中也仍然存在。

国际象棋之后,研究人员便把目标锁定在围棋上。国际象棋的搜索宽度大概是30,搜索深度大概是80,整个搜索空间大约为10^{50};而围棋的搜索宽度大概为250,搜索深度大概是150,搜索空间在10^{170}以上,比宇宙中的粒子数10^{80}还多。由于搜索空间太大,只依赖评估函数和剪枝搜索算法在有限的时间内无法完成对整个空间的搜索。因此,"深蓝"所使用的暴力穷举的搜索方法对于围棋则完全失效。很长时间以来,人们认为围棋是人工智能不可逾越的一道坎。人类为迈过"围棋"这道坎,足足准备了20年。

2 AlphaGo的进化之路

2016年1月26日,谷歌旗下的DeepMind团队在《Nature》杂志发表《Mastering the game of Go with deep neural networks and tree search》(《通过深度神经网络和树搜索来征服围棋》),从而揭开了围棋人机大战的历史性一页。该论文称,在2015年10月5—9日的比赛中,AlphaGo以5:0的比分战胜了欧洲围棋冠军樊麾(Fan Hui)[6]。这是围棋历史上机器第1次战胜职业围棋选手。为了进一步测试AlphaGo的性能,2016年3月,DeepMind团队向围棋世界冠军、韩国顶尖棋手李世石发起挑战。2016年3月9—15日,在韩国首尔举行的人机大战中,AlphaGo Lee以4:1的比分战胜了李世石。此后,DeepMind依据AlphaGo Lee与李世石的对战的经验,对系统做了进一步改进。

2016年12月29日起,一个名为"Master"的神秘网络棋手在几个知名围棋对战平台上轮番挑落中、日、韩围棋高手,并在2017年1月3日晚间战胜

中国顶级围棋棋手柯洁。此后,在击败古力使自己对人类的连胜纪录达到60:0后收手。正当人们对此神秘棋手进行揣测之时,DeepMind团队发表正式声明,宣称Master乃是AlphaGo Lee的升级版本。

2017年5月,在中国乌镇举行的人工智能峰会上,排名世界第一的围棋冠军柯洁挑战AlphaGo Master,最终以0:3落败。在比赛结束后的发布会上,DeepMind的负责人哈萨比斯(D. Hassabis)宣布AlphaGo"退役",即不再与人类棋手进行比赛,但申明团队仍旧会继续研究和发表相关的研究论文。

2017年10月,DeepMind团队公布进化最强版AlphaGo Zero[7],这个版本最大的特征是不再需要人类经验数据,用于训练的是机器自我对弈所产生的数据。在经过3天的训练后,AlphaGo Zero就可以战胜AlphaGo Lee,比分高达100:0;而经过40天的训练后,就以89:11的比分击败了AlphaGo Master。AlphaGo Zero的具体细节以论文的形式于2017年10月19日发表在《Nature》杂志上。

AlphaGo的成功,可归结于机器学习与人工神经网络相结合而产生的深度学习的应用。当然,谷歌的并行计算系统、TPU专用芯片以及大数据为它的成功提供了平台和物质基础。正如李开复所说:"深度学习、大规模计算和大数据三位一体。"[8]

2.1　人工神经网络与机器学习

1943年,神经科学家麦卡洛克(W. McColloch)和皮茨(W. Pitts)提出了模拟神经网络的理论,描述了人类神经沿网状结构传递和处理信息的理论模型。这一理论很快被计算机领域的研究者所借鉴,通过计算机模拟人的神经系统的工作模式来进行简单的模式识别和信息处理,从而开辟了人工神经网络(artificial neural networks)的研究领域。1957年,康奈尔大学的实验心理学家罗森布拉特(F. Rosenblatt)在计算机上实现了一种"感知机"的神经网络模型,并且证明了单层神经网络在处理线性可分的模式识别问题时可以收敛,从而使这一领域成为当时的热点研究领域之一。1965年,伊瓦赫年科(A. G. Ivakhnenko)提出了基于多层神经网络的机器学习模型,即现在所说的深度学习(deep learning)。正当人工神经网络研究大热之时,1969年,明斯基(M. Minsky)却断言感知机不能解决异或问题,罗森布拉特之前也意识到这一问题。这一论断使神经网络研究遭受了巨大的打击,并且也因此沉寂多年。直到1975年,在哈佛大学的沃波斯(P. Werbos)解决了这一难题之后,人工神经网络的发展才又逐渐进入了正轨。

一直以来,人工神经网络是机器学习的主要算法之一,主要用于计算机对图像、文字、语言等的识别。由于浅层的神经网络在实践中效果并不好,所以人们就逐渐就产生了用多层神经网络(图1)使计算机获得"学习"功能的想法。

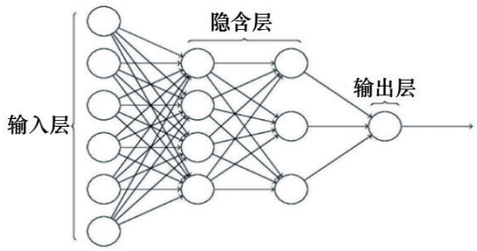


图1　多层神经网络示意

Fig. 1　Multi-layer neural networks

网络层数的增加将面临3个挑战:其一是在理论上无法解决由于网络层数增加而产生的问题;其二是层数越多,所需要的计算复杂度就越高,而计算机远远达不到处理深度神经网络的要求;其三是对复杂模型的训练需要海量数据,当时还没有大数据可用。因此,用大数据来训练复杂模型是深度学习的一个明确方向[9],等待的只是一个时机而已。

人工神经网络领域的泰斗辛顿(G. Hinton)是深度学习的推进者。2006年,他与合作者发表文章《A fast learning algorithm for deep belief nets》(《一种深度的置信网络的快速学习算法》)[10],提出了对深层神经网络进行训练的算法。该算法不但可以让计算机渐进地进行学习,而且学习的精确性会随着网络层数的增加而提高,这在客观上推动了无监督学习(unsupervised learning)的产生和发展。

2010年,谷歌开发了名为Google Brain的深度学习工具,将人工神经网络并行实现,即将一个大规模的模型训练的问题简化到同时能够分布到上万台服务器上训练的小问题,从而解决了深度学习

的基本技术问题。起初,谷歌大脑只使用了大量的CPU,后来又逐渐引入了GPU以及专用的TPU处理器。与此同时,大数据如火如荼的态势,使得基于深度网络的机器学习突破了前面所述的三大挑战,并在各领域都迅速发展,而围棋便是谷歌大脑的试金石。

2.2 AlphaGo的技术分析与演变

在数学上,“最优策略”和“判断局面”可以被量化成为函数$Q(s,a)$和$V(s)$。s表示局面状态,a表示落子动作。在关于强化学习的理论中,$Q(s,a)$被称为策略函数(policy function),$V(s)$被称作是局面函数或者评估函数(value function)。策略函数的用处在于衡量在局面s下执行a所能带来的价值;估值函数用于衡量局面s的价值,估值越大意味着在该落子动作下获胜的概率越高。因此,这两个函数可以用于模仿人类的下棋行为。人类在下棋时,首先凭借经验和“直觉”确定落子的若干方案(最优策略),这一行为客观上降低了“搜索宽度”,因为一些明显不好的方案不会被考虑进去;其次,对于每一个落子动作之后的情况,棋手也只能看到为数不多的几步(最顶尖的棋手在10步左右),并且以此为基础进行判断(判断盘面),这种行为在客观上降低了“搜索深度”。

AlphaGo在进化过程中,出现了4个典型的版本:AlphaGo Lee、AlphaGo Master、AlphaGo Zero及AlphaZero,以下将详述这些版本的设计思想和技术特征。

2.2.1 AlphaGo Lee

AlphaGo的设计思想就是模仿人类下棋的模式:用策略网络(policy network)来减小“搜索宽度”,即实现对人类“棋感”的模拟;用估值网络(value network)来减小“搜索深度”,从而模拟人类对盘面的综合判断;最后借助谷歌的技术优势——海量数据、并行计算以及GPU、TPU,通过训练最终获得远超人类棋手的棋力。

从技术上讲,AlphaGo在设计过程中,融合了蒙特卡洛树搜索算法(MCTS)、强化学习(RL)和深度神经网络(DNN)这3种目前人工智能领域最先进的技术。在具体设计中,蒙特卡洛树搜索为AlphaGo提供了一个基础框架;强化学习则用来提升AlphaGo的学习方法;深度神经网络则是用来拟合策略函数和估值函数的工具。这三大技术虽然在AlphaGo出现之前就已经成熟,但是谷歌借助于其巨大的计算能力(GPU、TPU、并行计算)以及海量数据,将三者有机结合,从而使AlphaGo获得了巨大的成功。

在设计思想上,AlphaGo设置有两个大脑(图2):一个是策略网络,另一个是估值网络,蒙特卡洛树搜索将这两个大脑整合在一起。通过这两个大脑,一方面来模拟人类下棋的“棋感”和大局观,另一方面来模拟人类对每一步棋的深思熟虑。因此,AlphaGo最终具备了在“直觉”基础上的“深思熟虑”,而这正是一种典型的“人类思维”处理复杂问题的方式,这为解决复杂决策智能的问题提供了一种工程技术框架[11]。

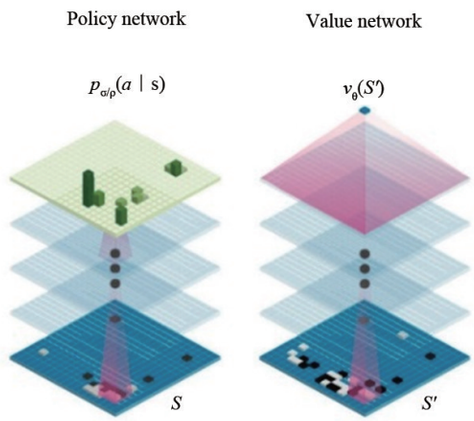


图2 两个大脑:“策略网络”和“估值网络”

Fig. 2 Two brains: Policy network and value network

具体方法上,第一搭建了一个具有13层的深层神经网络,并且用监督学习(supervised learning)的方式,用人类经验数据(KGS上的3000万盘棋局)对此网络进行训练,最终得到策略网络p_σ。具体训练方法是通过输入当前的棋盘数据,输出下一步落子位置的概率分布,从而预测落子位置。设计这个网络的目标不是为了赢棋,而是要从经验数据中归纳出最好的落子方法——“妙手”。这个网络对人类专家下棋预测的精准度达到了55.7%。一

方面,这个预测结果已经远超当时最先进的机器棋手(44.4%);另一方面,出现的不精准结果也不能全归于网络本身,因为人类棋手在落子时存在不可避免的"臭招"。

第二,虽然更庞大、复杂的神经网络能提高预测的精确度,但这也会拖慢网络评估的速度;此外,网络虽然能根据当前盘面,给出下一步落子的最好位置,但是它并不会"看棋",即不会给出后面的走法。为了解决这一问题,研究人员又通过人类棋谱训练了一个具有较少层数(双层)的神经网络——快速走子网络(rollout policy)p_π(图3)。这个网络虽然只能达到24.2%的预测精度,但与策略网络p_σ相比,下棋的速度更快——它只需2 μs,而p_σ需要3 ms,相当于快了1000多倍。之所以设置快速走子网络,原因是策略网络虽然精确,但是搜索速度慢,并且也不可能搜索到最后一步。而快速走子网络p_π的搜索结果(相同时间内搜索更深),可以让策略网络p_σ的实际搜索范围缩小,即实现对搜索树的剪枝。

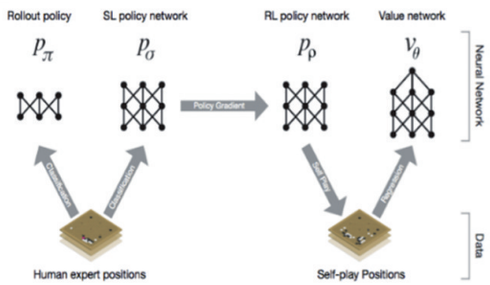


图3　网络训练、生成示意

Fig. 3　Neural network training pipeline and architecture

第三,通过强化学习来提高策略网络p_σ的棋力,得到策略网略p_ρ。p_ρ的结构与p_σ完全相同,其获得方法是,取p_σ作为1.0版本,通过左右手互博(自对弈),得到N个棋谱;接着用这N个棋谱对1.0版本进行训练,得到2.0版本;然后让1.0和2.0版本双方互博,再得到N个棋谱;再用这N个棋谱对2.0进行训练得到3.0。新的版本随机选择之前的版本进行互博,产生更新的版本,以此类推。通过n次的训练,最终得到p_ρ。整个过程中,自对弈数达到了3000万局。在实测中,p_ρ对最强的开源围棋软件Pachi的胜率达到了85%,而p_σ对Pachi的胜率只有11%[12]。可以看出,p_ρ是在人类经验的基础上,再依据机器互博的机器"经验",从而大幅度地提升了自己的棋力。数据显示,使用机器自对弈数据作为样本训练而成的版本,都具有较高的棋力,这也为谷歌最终完全放弃人类经验数据埋下了伏笔。

第四,通过机器经验(自对弈数据)对p_ρ进行训练,得到一个估值网络v_θ,这就是AlphaGo的第2个大脑,用于对盘面进行评估。估值网络的体系结构与策略网络近似,在功能上的不同在于它输出的不是一个概率分布空间,而是一个单一的预测结果。它会将无用的走法因其概率较低而剪枝,因此不再进一步搜索,从而极大地降低了搜索的深度。

第五,估值网络虽然通过剪枝减小了搜索深度,但是却不能给出最终的决策。最终的落子决策则是通过蒙特卡洛算法实现的。在采样不足的情况下,蒙特卡洛算法可以通过尽可能多次的随机采样,一步一步接近最优解。AlphaGo即是运用蒙特卡洛树搜索算法,对两个大脑即策略网络p_ρ与估值网络v_θ进行整合。蒙特卡洛树搜索的作用就是在模拟下棋的过程中对盘面进行评估。"Monte Carlo"一词源自于意大利语,有可疑、随机等意思。如果下棋的时候棋手"随机"落子,则必输无疑。因此,首先使用策略网络来预测人类的落子行为,即在AlphaGo的每一个落子动作执行之前,AlphaGo首先运行策略网络,从而获得一个人类棋手落子位置的概率分布;接着,蒙特卡洛搜索算法才以这个概率分布为基础进行"随机"。因此,正因二者的有效结合,才使得AlphaGo在减小搜索空间的基础上,得到了赢棋概率最高的落子动作。

从对AlphaGo的技术分析可以看出,首先,该系统的成功离不开3个方面:数据、硬件和算法,三方互相依赖,缺一不可。数据方面,AlphaGo将围棋盘面的一个状态s抽象为19×19的网格图像,并且抽取出48个特征量来表征这一状态。因此,每一个状态s是一个19×19×48的图像。在整个训练中,先后用到了KGS的3000万盘棋局以及自对弈产生的3000万盘棋。对于一个围棋学习者来说,想达到顶级水平所需完成的盘数大概为几万盘。

因此,AlphaGo所用的数据是海量的大数据。在硬件方面,谷歌强大的硬件系统为训练深度神经网络提供了基础:与李世石比赛的AlphaGo Lee使用了40个搜索线程、48个CPU、8个GPU。而最强的分布式的AlphaGo版本,利用了多台电脑,40个搜索线程、1202个CPU、176个GPU。在算法上,AlphaGo使用了机器学习中的监督学习、加强学习和蒙特卡洛搜索算法。这些算法虽然都早已有之,但是在谷歌大数据以及强大的计算技术的支持下,显示出了巨大的威力[12]。其次,依靠人类经验,机器最多只能达到人类顶尖水平,而要超越人类,就需要摒弃人类的经验。相对于人类数据,机器经验数据已经是较优数据。由机器经验数据训练而来的value network可以达到专业5段水平,远高于通过人类经验训练而来的rollouts和policy network(图4)。

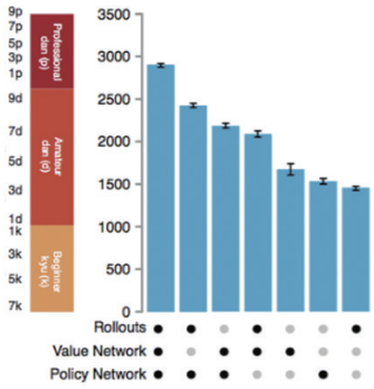


图4 3种网络以不同方式组合后的棋力水平

Fig. 4 Performance for different combinations of components

2.2.2 AlphaGo Master

DeepMind虽然没有发表关于AlphaGo Master的文章,但可以结合它的表现以及DeepMind团队负责人哈萨比斯在乌镇的演讲,大致梳理出一些关键的技术和思想。

与AlphaGo Lee相比,AlphaGo Master在下棋的表现上有以下典型的特点。

第一,落子速度更快,发挥更稳定。上线以来,除了由网络中断而出现的自判和局以外,接连战胜世界顶级高手,对人类的连胜达到60:0。

第二,布局与落子方法极具创造性,并且形成了许多人类高手没有见过的布局与落子方法,对人类极具启发性,使人类对棋理产生了重新认识。柯洁在观看AlphaGo Master在线的部分表现后表示:“从来没见过这样的招法,围棋还能这么下?”“看AlphaGo Master的招法,等于说以前学的围棋都是错误的,原来学棋的时候要被骂的招法现在AlphaGo Master都下出来了。”

第三,AlphaGo Master基本上都是在中盘就已经确定了绝对优势。由此看来,人类棋手经过几千年时间所归纳总结的经验知识与AlphaGo Master相比,显然不在一个层次。此外,最为关键的是,AlphaGo Master似乎还有上升空间。

在设计思想上,DeepMind对AlphaGo Lee的两个大脑即策略网络和估值网络做了进一步改造和优化。从其超越人类围棋知识的表现来说,已经与AlphaGo Lee有极大的不同,并且可以肯定的是对人类经验数据的依赖变得更少。在AlphaGo Lee输掉的第4局中,李世石在第78手下出的“神之一手”,导致机器如同人类棋手一样产生由于缺乏经验而产生的“慌乱”,并且接连出现欠考虑的“臭招”,最终落败,这说明人类经验数据对机器下棋策略的某种限制。DeepMind的优化思路是用2个AlphaGo Lee自我对弈,用对弈得到的机器经验数据再进行强化学习,从而得到新的策略网络和估值网络,然后再将2个网络整合,得到一个加强的版本AlphaGo Master。由于这个版本的训练数据完全是自我对战的高质量机器数据,因此由其所训练得到的网络更强大,并且也进一步缩小了树搜索的搜索空间。数据表明,在硬件系统上,由于AlphaGo Master需要的计算量是AlphaGo Lee的1/10,与柯洁对战的AlphaGo Master实际上已经实现了单机运行,并且只用到了4个TPU。尽管如此,它的棋力就已经远超了AlphaGo lee。

由于在训练过程中完全使用了机器自对弈数据,棋力远远超过了先前基于大量人类经验数据训练而来AlphaGo Lee,这进一步证明了人类经验的局限性。而在硬件系统上的低要求说明人类数据训练而来的策略网络是拖慢系统速度的主因。因

此,完全放弃人类经验数据就成为DeepMind的必然选择。

2.2.3　AlphaGo Zero

2017年10月19日,DeepMind团队在《Nature》发表文章《Mastering the game of Go without human knowledge》(《不需要人类知识的围棋游戏》),论文指出:“由之前AlphaGo的训练和对弈经验可以看出,人工智能的许多进展都是通过监督学习而取得的,即通过专家数据集来训练系统,以模拟人类专家的决策。但是专家数据集通常是昂贵的、不可靠的或根本是不可用的。即使是可靠的数据集,但它们也可能对以这种方式训练得到的系统的性能造成限制。”[13]因此,AlphaGo Zero最大的亮点就是完全放弃围棋的人类经验知识,以围棋的规则为基础框架,通过自对弈从而得到机器自身关于围棋的知识。其基本思路和技术特征可以归结为:第一,放弃之前所用的卷积神经网络,而是选用残差神经网络,这个网络比AlphaGo Master所用的卷积网络更为复杂,它包含40个隐含层,比AlphaGo Master多一倍;第二,只用一个大脑,而不是如同之前的版本由卷积神经网络训练出的两个大脑——策略网络和估值网络;第三,基于围棋的基本规则,以最为直接和简单的黑白子为输入特征量,进行无监督的加强学习;第四,完全放弃围棋领域的人类经验数据,机器从零开始不断地左右互博(4900万盘),以寻找和归纳围棋知识;第五,只使用最简单的MCTS树搜索,并依赖单一的神经网络来预测落子位置和评估盘面;第六,为了提高学习的速度和保证精确、稳定的学习过程,开发和使用了一种新的加强学习算法。

最终训练成的AlphaGo Zero能在具有4个TPU的单机上运行(训练过程中使用了CPU、GPU和TPU),并且在训练3 h后棋力就达到了AlphaGo Lee的水平,训练40 d就能达到AlphaGo Master的水平(图5)。可以看出,在摒弃人类经验数据、改变方法以及引入新算法后,不但使AlphaGo的棋力大大提高,而且也降低了运行过程中的能耗,极大地提高了效率。

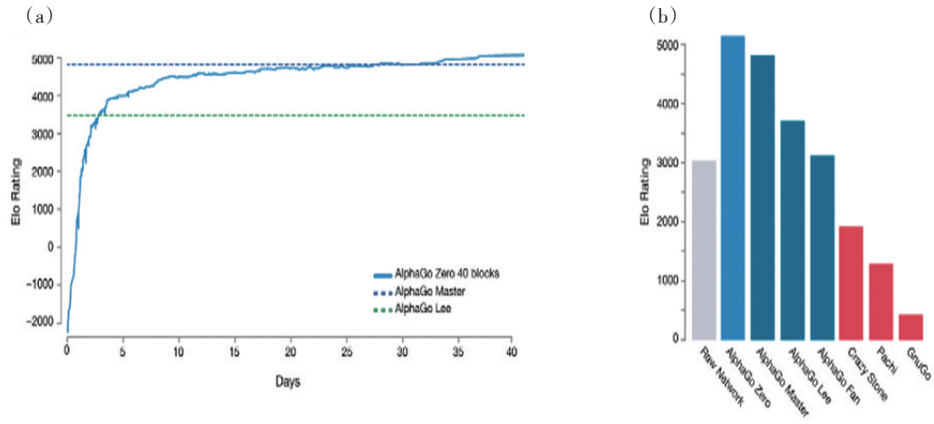


图5　不同版本的AlphaGo的训练时间与棋力表现

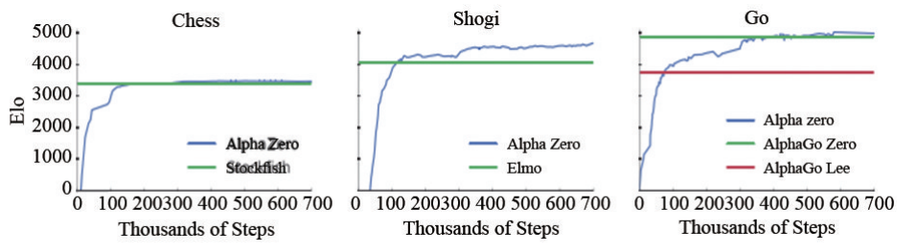
Fig. 5　Performance of each program on an Elo scale

自围棋诞生以来的几百万年的时间里,人们通过无数次的对弈游戏,积累了大量的围棋知识、定式和书籍。而AlphaGo Zero只通过3 d的训练,从围棋“儿童”达到围棋“超人”的水平。棋力不但远远超越了人类水平,同时还发现了人类未发现的新的知识、定式等。

2.2.4　AlphaZero

2018年12月,《Science》杂志发表论文《A general reinforcement learning algorithm that masters chess, shogi and Go through self-play》(《用通用强化学习算法自我对弈,掌握国际象棋、将棋和围棋》)。论文揭示,DeepMind依据之前的经验,采用新算法开发了单一系统AlphaZero,这套系统竟然在短期的自我学习中,成功地实现了对国际象棋、日本将棋及围棋目前最强智能系统的完胜。论文揭示:AlphaZero仅用4 h的自我学习,就超越了目

前最强的国际象棋智能系统Stockfish;仅用2 h的自我学习就超越了日本将棋的最强智能系统Elmo;仅用8 h就战胜了围棋最强智能系统AlphaGo Zero(图6)[14]。



(a) Chess (b) Shogi (c) Go

图6 AlphaZero在训练700000步时所达到的水平

Fig. 6 Training AlphaZero for 700,000 steps

AlphaZero的出现标志着人类在信息完全博弈领域(至少是棋类游戏)实现通用智能系统的关键性进展。国际象棋大师卡斯帕罗夫在应邀参加AlphaZero与国际象棋系统的对战后感慨道:“我真的不能掩饰我自己的满足感,它极具活力,就如同我一样!”[15]

综上可以看出,在人类自认为所擅长的棋类游戏领域,机器能在不需要人类经验、专家知识的情况下,短时间内同时掌握多种游戏技能,并且实现对人类的全面超越;而在机器擅长的方面,“人的因素成为了一种导致错误的诱因。人类,仅凭其迟缓的反应时间和高度的易疲劳性,根本无法与计算机和高速设备相匹敌”[16]。面对机器的迅速进化和崛起,人类该怎样面对呢?

3 如何应对机器的崛起

纵观机器下棋的短短的20多年的历史,尤其是AlphaGo不到3年时间的“进化史”,充分说明,在围棋这个领域,受限于人类自身认知能力的局限,人类几千年积累下来的经验数据较之于机器在3 d内形成的经验数据,已经不是最优数据。正如柯洁在输掉比赛之后曾表示:“围棋从产生到现在已经经历了几千年的历史,但是AlphaGo却向人类表明,人类可能还没有揭开围棋的表皮。”[17]也如DeepMind的论文中所指出的:“专家数据集通常是昂贵的、不可靠的或根本不可用的。即使可靠的数据集是可用的,也可能对以这种方式训练的系统的性能造成限制。”[13]放弃对人类经验的依赖,深度强化学习算法也许能被广泛应用到其他复杂领域,尤其是信息不完全的复杂系统,如天气、医疗等[18]。

面对智能系统的迅速崛起,人类不能总是陷入塞尔(J. R. Searle)“中文屋”论证的泥淖,即:声称机器虽然能战胜人类,但它不懂得下棋,机器有的仅仅是“一串串无意义的符号”。其实,马丁·戴维斯(M. Davis)对这一论证早就有过精辟的反驳:“塞尔强调了‘深蓝’不‘知道’任何东西,而富有专业知识的工程师却有可能声称,‘深蓝’的确知道各种东西,例如它知道能将给定方格中的象移动到哪几个方格中去,这完全取决于‘知道’是什么意思。”[19]毕竟目前的机器与以前的机器有了巨大的不同。机器在进化中不断地蜕变,它已经不仅仅是对人类感官系统的放大,也不仅仅是作为人类认知和行动的辅助系统,“他”已经逐渐拥有了自己独立的价值与生命[18]。即使机器不会像人类一样去“理解”世界,但是在某些方面却能比人类做得更好,因此,人类要谦逊地接受和面对[20]。

此外要说明的是,AlphaGo战胜人类世界冠军,只说明它在围棋上比人类做得更好,它并没有全面攻克围棋。因为AlphaGo是以胜利、赢得比赛的实用主义哲学为唯一目标,而不是以追求必胜策略或最优理论的理性主义为目标。要想真正攻克围棋,路还很长。就拿跳棋游戏来说,由哈佛大学的舍佛(J. Schaeffer)团队设计的Chinook跳棋程序

于1994年就战胜了当时的跳棋冠军丁斯利(M. Tinsley),但直到2007年,舍佛团队才从理论上证明,对于跳棋,“只要对弈双方不犯错,最终都是和棋”[2]。从这种意义上讲,战胜并非是完全攻克,这本质上是由于其核心——人工神经网络——的认识论本质所决定的,因为神经网络作为复杂网络系统,通过搭建神经元之间的网络关系,模拟人脑的结构和功能,是对大脑信息处理方式的简化、抽象和模拟,输入与输出之间的对偶是通过复杂的参数调整,因此学习目标的达成并非是基于因果律的。在完全信息的棋类游戏中,由于本质上每一步棋有固定的走法,如果算力足够强大,则完全可以“计算”每一步棋的最优走法。可以说,完全信息的游戏本质上就是计算,只不过,“深蓝”依赖的是高速芯片,AlphaGo则依赖“谷歌大脑”提供的算力以及新算法对搜索空间的剪枝。也就是说,即使仅靠算力最强的谷歌大脑,都无法穷尽围棋的所有可能性。如果假设有更强的算力支撑的话,用最简单的暴力穷举法就完全可以战胜人类冠军。比如按照目前的量子计算机的理论,如果能实现600个量子位的量子计算机,其计算能力就能达到10^{180},这个结果很显然超过了围棋的局面数。

基于以上的讨论,可以做如下总结:首先,从IBM的“深蓝”,到谷歌的AlphaGo,再到AlphaGo Zero,经历了对人类经验的重新审视以及对机器自身“经验”的新认识。也就是说,对于信息完全的博弈系统,依据强大的算力,机器可以做得比人更好,机器获得的“经验”比人类经验更优。其次,新的算法的开发不仅可以降低机器对算力的依赖,而且可以摆脱人类经验的束缚。机器可以依靠基础规则,通过不断的自我对弈,达到远超人类经验、知识的水平。AlphaGo的搜索深度比深蓝系统的搜索位置少了很多,这主要归功于深度神经网络算法的开发。再次,AlphaGo对于人类最大的启示在于:放弃对人类经验的依赖,深度强化学习算法也许能被广泛应用到其他复杂领域,尤其是信息不完全的复杂系统,如天气、医疗等。这也正是DeepMind团队的初衷和目标绝不在于攻克围棋的原因。围棋,只是他们的一个试金石或者小战场,医疗、癌症、天气预测等缺乏完全信息的复杂系统才是他们的终极目标。DeepMind的负责人哈萨比斯在多个场合就说过,他们的目标在于将在游戏中证明过的技术,用来解决医疗等更为复杂的问题。这些问题,对于最聪明的人都是无可奈何的,人工智能是一个解决这些复杂问题的潜在模式。“我们发明AlphaGo并以此来探索围棋的奥秘,正如科学家用哈勃望远镜来探索宇宙的奥秘一样。因此,AlphaGo的发明,并不是为了战胜人类。与人类进行比赛,是为了测试我们的智能算法,因此它只是手段,而不是目的。这些有效的算法应用到真实世界,并为人类社会提供服务才是我们的终极目标。”[21]

聂卫平说:“人类可以向AlphaGo学习!”2008年《Nature》大数据专刊中讨论的主题是“人类从谷歌能学到什么”。面对快速发展的智能系统,人类需要从智能系统中学习什么,这正是人类需要深思的。

参考文献(References)

[1] Shannon C E. Programming a computer for playing chess [J]. Philosophical Magazine, 1950, 41(314): 256-275.

[2] 尼克. 人工智能简史[M]. 北京:人民邮电出版社, 2017.
Ni Ke. A Brief History of Artificial Intelligence[M]. Beijing: Posts & Telecom Press, 2017.

[3] Newborn M. 旷世之战——IBM深蓝夺冠之路[M]. 邵谦谦, 译. 北京: 清华大学出版社, 2004.
Newborn M. Deep Blue—An artificial intelligence milestone[M]. Shao Qianqian, trans. Beijing: Tsinghua University, 2007.

[4] Hsu F H. IBM′s Deep Blue chess grandmaster chips[J]. IEEE Computer Society Press, 1999, 19(2): 70-81.

[5] 吴岸城. 神经网络与深度学习[M]. 北京: 电子工业出版社, 2016.
Wu Ancheng. Neural network and deep learning[M]. Beijing: Publishing House of Electronics Industry, 2016.

[6] Silver D, H A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. Nature, 2016, 529(7587): 484-489.

[7] Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of Go without human knowledge[J]. Nature, 2017, 550(7676): 354-359.

[8] 李开复, 王咏刚. 人工智能[M]. 北京: 文化发展出版社, 2017.
Li Kaifu, Wang Yonggang. Artificial Intelligence[M]. Beijing: Cultural Development Press, 2017.

[9] 吴军. 智能时代: 大数据与智能革命重新定义未来[M].

北京: 中信出版集团, 2016.
Wu Jun. The age of intelligence: Big data and the intelligent revolution redefine the future[M]. Beijing: China CITIC Press, 2016.
[10] Hinton G E, Osindero S, Teh Y. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7): 1527-1554.
[11] 陶九阳, 吴琳, 胡晓峰. AlphaGo技术原理分析及人工智能军事应用展望[J]. 指挥与控制学报, 2016, 2(2): 114-120.
Tao Jiuyang, Wu Lin, Hu Xiaofeng. Principle analysis on AlphaGo and perspective in milltary[J]. Application of Artificial Intelligence, 2016, 2(2): 114-120.
[12] Silver D, H A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. Nature, 2016, 529(7587): 484-489.
[13] Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of Go without human knowledge[J]. Nature, 2017, 550(7676): 354-359.
[14] Silver D, Hubert T, et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play[J]. Science, 2018, 362(6419): 1087-1118.
[15] AlphaZero: Shedding new light on the grand games of chess, shogi and Go[EB/OL]. [2019-07-05]. https://deepmind.com/blog/alphazero-shedding-new-light-grand-games-chess-shogi-and-go.
[16] Sterken C, Manfroid J. Astronomical photometry: A guide [M]. Springer Science & Business Media, 1992.
[17] AlphaGo之父: 关于围棋, 人类3000年来犯了一个错[EB/OL]. [2019-07-05]. https://www.thepaper.cn/newsDetail_forward_1660773.
Father of AlphaGo: A mistake about go humans made for 3000 years [EB/OL]. [2019-07-05]. https://www.thepaper.cn/newsDetail_forward_1660773.
[18] 董春雨, 薛永红. 机器认识论何以可能[J]. 自然辩证法研究, 2019, 35(8): 3-10.
Dong Chunyu, Xue Yonghong. Why is machine epistemology possible?[J]. Studies in Dialectics of Nature, 2019, 35(8): 3-10.
[19] 马丁·戴维斯. 逻辑的引擎[M]. 张卜天, 译. 长沙: 湖南科学技术出版社, 2001.
Martin Davis. Engines of Logic[M]. Zhang Butian, trans. Changsha: Hunan Science and Technology Press, 2001.
[20] Alvarado R, Humphreys P. Big data, thick mediation, and representational opacity[J]. New Literary History, 2017, 48(4): 729-749.
[21] 哈萨比斯在剑桥大学的演讲"超越人类认知的极限"[EB/OL]. [2019-07-05]. http://scholarsupdate.hi2net.com/news.asp?NewsID=22161.
Demis Hassabis's talk at Cambridge University about "Exploring the frontiers of knowledge"[EB/OL]. [2019-07-05]. http://scholarsupdate.hi2net.com/news.asp?NewsID=22161.

Brief history and enlightenment of machine chess: From "Deep Blue" to AlphaZero

XUE Yonghong[1,2], WANG Hongpeng[3]

1. College of Science, North China University of Science and Technology, Beijing 101601, China
2. College of Philosophy, Beijing Normal University, Beijing 100875, China
3. China Science and Technology Museum, Beijing 100012, China

Abstract Taking the historical evolution as the main line, this paper combs and summarizes "Deep Blue" and AlphaGo from the aspects of design philosophy and technical features. Relied on human experience of chess, "Deep Blue" achieved transcendence with humans by means of computational power and algorithms. Twenty years later, AlphaGo, although its original version was also successful by using human experience, and its evolution revealed an important fact that human experience has its limitation. AlphaZero, which gives up human experience and adopts machine self-playing experience, convincingly defeated a world champion program in the games of chess and shogi (Japanese chess) as well as Go. It is clear that in chess games, limited by human cognition ability, experience accumulated by human beings for thousands of years is no longer superior to the "experience" formed by machines in a short term. Machines can do better than humans with the help of their own "experience", supported by enormous computing power and ever-improving algorithms. In the future, machines that "give up human experience and rely on their own experience" will likely make breakthroughs in more complex areas.

Keywords machine chess; Deep Blue; AlphaGo; AlphaZero

（责任编辑 刘志远）