AlphaGo的突破与兵棋推演的挑战

胡晓峰[1],贺筱媛[1],陶九阳[1,2]

1. 国防大学信息作战与指挥训练教研部,北京 100091
2. 陆军工程大学指挥信息系统学院,南京 210007

摘要 概述了AlphaGo的原理、方法创新、技术突破和在认识论上的意义。分析了兵棋推演面临的瓶颈,指出了作战智能态势认知是亟需突破的关键环节。提出了解决作战态势智能认知的实现途径。展望了"人机智能"为兵棋推演带来的新机遇。

关键词 AlphaGo;深度学习;兵棋推演;态势认知

2016年3月,谷歌公司AlphaGo与围棋人工智能程序李世乭进行的"人机围棋大战"引起了全世界的关注,自然而然地使人将AlphaGo与兵棋推演联系起来。那么,AlphaGo究竟突破了什么?它能否直接用于兵棋推演?对兵棋系统研发能带来哪些帮助?有哪些难题亟需突破?本文将就以上问题展开论述。

1 AlphaGo带来了什么

1.1 AlphaGo取得的4个突破

纵观历史,"人机大战"已有多次,最具代表性的主要有3次:第一次是1997年IBM公司的"深蓝",以及后来"更深的蓝"先后击败国际象棋大师卡斯帕罗夫[1];第二次是2011年IBM公司的问答机器人"沃森",在美国《危险边缘》智力问答竞赛节目中大胜人类冠军詹宁斯,得到的奖金数超过了第一名和第二名的总和;第三次就是2016年的AlphaGo[2]与李世乭的围棋大战,在5场对弈中,以4:1的战绩战胜李世乭,令全世界哗然,宣告人类在一个引以为傲的智能高地上再次败北。AlphaGo之后,人机大战在各个领域出现井喷态势。2016年6月,人工智能飞行员Alpha AI(阿尔法鹰)战胜了美国空军著名战术专家李上校;2016年8月,卡耐基梅隆大学的Mayhem机器人战队经过95轮战后,战胜了所有人类战队,夺得美国国防高级研究计划局(DARPA)第24届网络挑战大赛(CGC)冠军。2017年初,AlphaGo化名"Master"(大师),在著名围棋对弈网站先后战胜世界围棋冠军15名,豪取60连胜;2月,卡耐基梅隆大学开发的人工智能系统Libratus在人机德州扑克大战中击败了人类顶级职业玩家;5月,AlphaGo再次以3:0的战绩战胜当今围棋排名第一的柯洁。AlphaGo的成功,点燃了人机大战的熊熊烈火。

AlphaGo之所以带给人们如此大的震撼,最重要的原因是其在"方法"上取得的突破,被认为是智能技术特别是机器智能技术进步的重要里程碑。博弈问题的核心,是解决巨量解空间中的评估选择问题[2]。围棋有19×19个格子,理论上其落子的选择为3^{361}。据估算,即使去掉不合理之处,可选择的数目也有170位之多 。在如此大的空间中找到好的落子方案,对当前已有的任何一种搜索算法来说,都无法在有限时间内完成,而AlphaGo最终较好地解决了这个问题。AlphaGo取得的重要突破主要表现在以下4个方面(图1)。

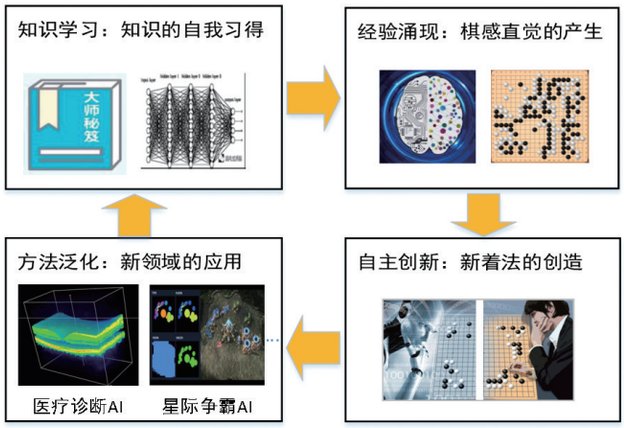

图1 AlphaGo展现出的技术突破

Fig. 1 Technological breakthroughs of AlphaGo

1)自学习能力。AlphaGo的对弈知识是通过深度学习方法自己掌握的,而不是像"深蓝"那样编在程序里,也不像"沃森"是通过"读书"来建立知识网络,它是通过大量棋谱和自我对弈完成的。尽管这个能力目前还很初级,但却展现了极好前景,使长期困扰人们的人工智能自学习问题有了解决

收稿日期:2016-09-06;修回日期:2017-06-18

基金项目:军民共用重大研究计划联合基金项目(U1435218);国家自然科学基金项目(61174156,61273189,61174035,61374179,61403400,61403401)

作者简介:胡晓峰,教授,研究方向为战争模拟、军事运筹和军事信息系统工程,电子信箱:xfhu@vip.sina.com

引用格式:胡晓峰, 贺筱媛, 陶九阳. AlphaGo的突破与兵棋推演的挑战[J]. 科技导报, 2017, 35(21): 49-60; doi: 10.3981/j.issn.1000-7857.2017.21.006

的可能。这种深度学习的能力,使得AlphaGo能不断学习进化,产生了很强的适应性,而适应性造就了复杂性[4],复杂自适应性又是智能演化最普遍的途径。

2)捕捉经验的能力。找到了一种捕捉围棋高手经验,即“棋感直觉”的方法。所谓棋感,就是通过训练得到的直觉,“只可意会,不可言传”。AlphaGo通过深度学习产生的策略网络(或称走棋网络),在对抗过程中可以实现局部着法的优化;通过增强学习方法生成的估值网络,实现对全局不间断的评估,用于判定每一步棋对全局棋胜负的影响。此外,还可以通过快速走子算法和蒙特卡洛树搜索机制,加快走棋速度,实现对弈质量和速度保证的合理折衷。这些技术使得计算机初步具备了既可以考虑局部得失,又可以考虑全局整体胜负的能力。而这种全局性“直觉”平衡能力,正是过去人们认为是人类独有、计算机难以做到的。

3)发现创新能力。发现了人类没有的围棋着法,初步展示了机器发现“新事物”的“创造性”。在五番棋的对抗过程中,从观战的超一流棋手讨论和反应可以看出,AlphaGo的着法有些超出了他们的预料,但事后评估又认为是好棋。这意味着AlphaGo的增强学习算法,甚至可以从大数据中发现人类千百年来还未发现的规律和知识,为人类扩展自己的知识体系开辟了新的认知通道。有人认为,AlphaGo的围棋水平已经达到了超一流的“十三段”,而人类最高才十段。所以“它可能比我们更接近围棋之神”,具备了超出人类对围棋博弈规律的理解能力。而人类也可以通过向计算机学习围棋,进一步加深对围棋规律的理解。

4)方法具有通用性。这与很多其他博弈程序非常不同,通用性意味着对解决其他问题极具参考价值。AlphaGo运用的方法,实际上是一种解决复杂决策问题的通用框架,而不仅是围棋领域的独门秘籍。自己学习的能力,使得计算机有了进化的可能,通用性则使其不再局限于围棋领域。AlphaGo的设计者曾声称,其下一步的目标是“星际争霸”,这是一个比较复杂的战争策略游戏,与实际的战争决策非常接近,说明这种技术框架具有广阔的应用前景[5]。

1.2 神经元网络与深度学习方法

AlphaGo之所以能获得棋感,其创新技术的核心就是建立在神经元网络基础上的深度增强学习方法。即通过建立神经元网络,并且尽量扩大学习的样本数,来理解概念、捕捉棋感。

神经元网络并不是一种新技术,其本质是对人脑神经网络的模拟,自20世纪80年代提出后就被广泛应用[6-8]。一个神经元网络通常由输入层、隐含层和输出层组成,隐含层类似于人类大脑皮层,隐含层越多,神经元网络的能力可能就越强。大脑皮层的厚度与人的智商正相关,大脑皮层越厚的人也极可能越聪明。然而,受计算能力和神经元网络训练算法复杂度的限制,长期以来,人们实际使用的神经元网络大多都只有一个隐含层。近几年,随着计算能力的提高和算法的不断优化及创新,具有多个隐含层的神经元网络开始大量涌现。这种具有多个隐含层的神经元网络被称为深度神经网络,而其构建方法就是深度学习[9-12]。

本质上说,深度学习是一种通过对大量样本的学习,形成对事物特征的提取和分类的方法。AlphaGo就是将围棋的每一个盘面用49个特征来描述[2],每一个棋子又有49个特征,并将这些特征通过多层连接构成一个神经元网络。神经元网络各层之间如何连接,就是对神经元网络的学习训练过程。实际上就是利用大量棋谱对神经元网络进行逐步改进,经过几十亿次的改进后,它就能逐渐理解人为什么这样下棋,在隐含层中不断对下棋的经验进行“总结”和“提炼”,从而获得人类的“棋感”。

深度学习方法可以概括为:大数据+高性能计算+神经元网络算法(图2)。也就是说,深度学习模型的建立和优化依赖于大量的样本数据,没有数据就无法训练深度学习模型;高性能计算尤其是基于图形处理器(GPU)的并行计算,可以极大地缩短模型进行大数据处理和运算调优的时间;同时,优化的神经元网络算法也能帮助模型提高计算效率。因此,深度学习本质上是一系列反复训练调优的神经元网络,其训练过程实际上就是整合理解的过程。

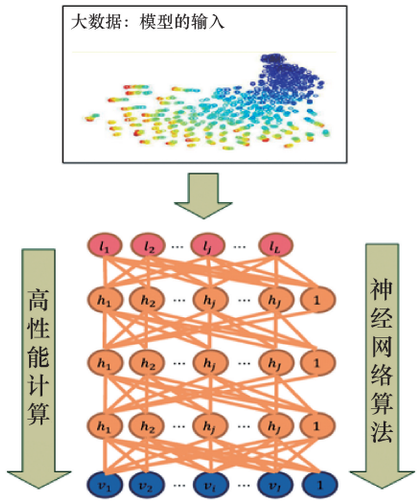

图2 深度学习模型原理

Fig. 2 Principle of deep learning

因此,通过深度学习方法得到的认知,实际上是一种非常接近于人类的认知方式。以往人类主要有两种认知方式:第一种是通过抽象化方式学习,例如学习牛顿定律、运动方程等,这些都是可以用非常形式化的方式进行表达的知识;第二种就是通过日常生活不断积累经验的方式学习,例如人类在玩接抛球游戏时,大脑里并没有一堆类似抛物线的公式,但却能稳稳地接住球,这种能力主要依赖于经验。经验往往需要通过大量类似场景的训练来获得,这种经验习得的方法并非人类所独有。例如,通过反复训练,最终小狗能够很好地接住飞盘,这就是在其大脑神经网络中形成了对这种场景的经验认知。

通过捕捉经验来学习是最常用、最复杂,但也最直接的学习过程,是一种普遍存在于高等生物中的认知方式。许多复杂事物因为找不到对应的因果关系、无法进行形式化表达,只能依据“棋感”、“经验”等所谓“只可意会,不可言传”的东西。因此,这学习需要过程,结果也可能有对有错,但它却正好使人们能够达成“捕捉经验”的结果。所以,从方法论角度来看,这也是一种更符合复杂系统本质的认知方式。

1.3 AlphaGo的原理与方法创新

AlphaGo主要依赖以下4个功能模块的融合和协同运行,来实现围棋对弈的智能决策。

1)策略网络(policy network)。主要功能是通过学习前人棋谱来获取走子经验,预测下一步棋的走子策略,并且可以通过反复自我对弈不断进化,从而实现版本升级和能力提升(图3(a))。策略网络又分为有监督学习策略网络(supervised learning policy network,SL策略网络)、快速走子策略(rollout policy)和增强学习策略网络(reinforcement learning policy network,RL策略网络)[2]。SL策略网络和RL策略网络都是一个13层的卷积神经网络[13-15],它们的输入为当前的盘面,输出是下一步棋盘上的落子概率,也就是可以得到下一步最有可能落子的位置。SL策略网络是根据人类高手的棋谱经验训练出来的,而RL策略网络是在SL策略网络基础上,通过自我对弈不断进化更高级策略的网络版本。

2)估值网络(value network)。主要功能是通过大量的自我博弈,完成对整个棋局胜负的判定预测(图3(b))。估值网络是一个13层的卷积神经网络,输入一个盘面,输出在这个盘面下赢棋的概率,完成对整个棋局胜负的判定预测。图4显示了AlphaGo在和李世乭下第一局棋时预测的实时胜负曲线。通过曲线可以看出,中盘以后,AlphaGo认为自己每一步都是领先的,说明它对整体形势的把握比较准确的。

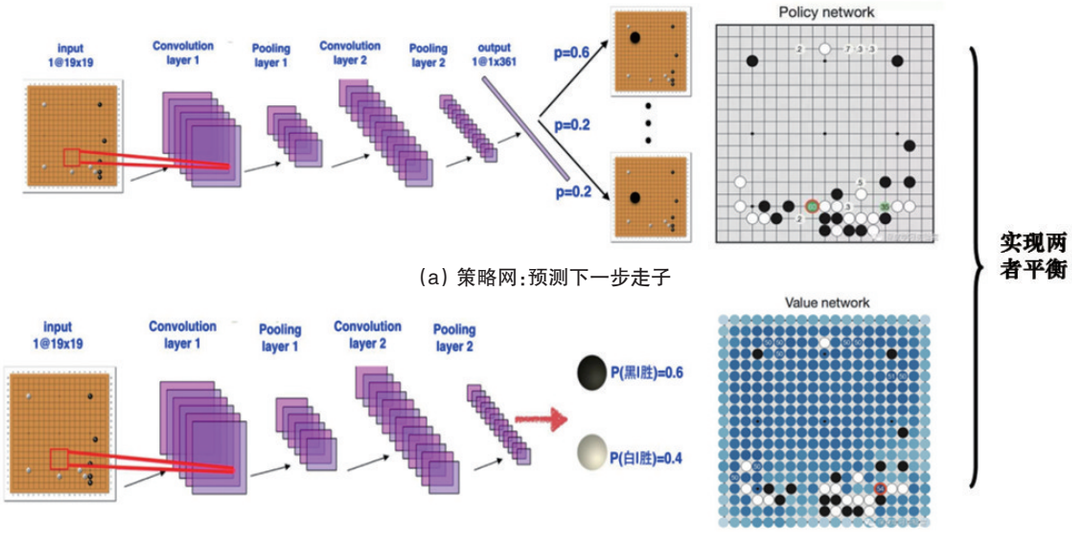

图3 AlphaGo利用策略网和估值网实现棋感与直觉的平衡

Fig. 3 AlphaGo achieved mix-strategies with policy network and value network

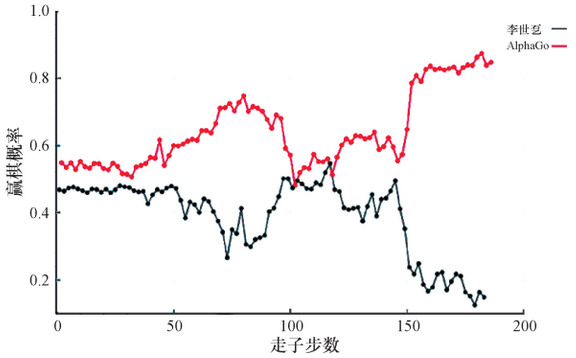

图4 AlphaGo估值网络预测的实时胜率曲线

Fig. 4 Win-rate prediction by AlphaGo's value network

3)快速走子(fast rollout)。主要功能是加快走棋速度。快速走子采用局部特征匹配与线性回归相结合的方法,通过剪枝来提高快速走子速度(图5),类似于深蓝的暴力搜索方法。其功能与SL策略网络类似,但结构是一个线性模型,比SL策略网络的卷积神经网络简单得多。

4)蒙特卡洛树搜索(MCTS)。主要功能是搜索计算后续步的获胜概率(图6[2])。相当于总控,控制对前3个算法的选择,完成对策略空间的搜索,确定出最终的落子方案[16-17]。为提高运算速度,计算可并行用GPU完成。

在上述4种算法中,AlphaGo用策略网络和估值网络两种方法的结合,解决了局部优化与全局平衡的问题(图3);快速走子只考虑局部的着法,但速度比策略网络快了大概1000

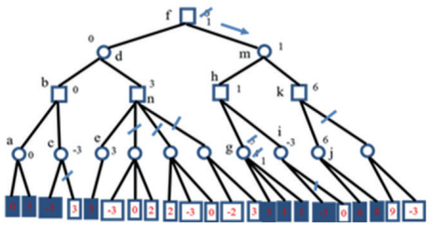

图5 AlphaGo通过剪枝提高快速走子速度

Fig. 5 Alpha-Beta pruning to speed up Rollout in AlphaGo

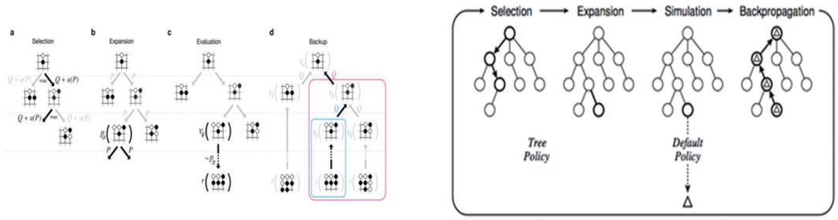

图6 利用蒙特卡洛树搜索计算后续步的获胜概率

Fig. 6 Monte Carlo tree search in AlphaGo

倍;蒙特卡洛树搜索结合前三者的优势,探索胜率高的着法,最终得到落子的点。

通过算法分析可知,AlphaGo利用策略网络模拟了人类在思维广度上的直觉估计能力,利用蒙特卡洛树搜索方法融合估值网络和快速走子,模拟人类在思维深度上的推演估算能力,从而综合广度和深度上的认知结果得出最优落子方案[18]。此外,最近披露的消息指出,AlphaGo并没有时间观念,其时间是由人控制的,只要人不喊停,它就在某一步棋上不断进行蒙特卡洛树搜索运算,给它的时间越多,模拟运算的次数越多,最终得到的结果就越精确。从认识论的角度看,AlphaGo的这种自我学习能力与人类智能十分相似,虽然这些神经元网络不能确保取胜,但可以确保胜率更高。

策略网络和估值网络的核心都是神经元网络,神经元网络的产生过程实际上是一个学习的过程,要教会它下棋,必须用数据进行训练。为此,AlphaGo收集了16万局人类高手的棋谱数据,每一局大约有200个盘面,共拆分为3000万手盘面(一步算一个盘面)进行训练。用估值网络判断整个棋局胜负时,由于16万盘棋每盘都只有一个胜负,AlphaGo又需要深度增强学习[2,19-21]技术进行自我博弈,所以又以半随机方式生成了3000万盘棋谱解决估值网络的训练问题。这是一种模拟人类智能非常有创意的一种想法,与"深蓝"的方法和"沃森"的智能问答有本质的不同。"深蓝"是按规则遍历各种可能的着法,即"暴力搜索",卡斯帕罗夫能向前深度搜索10步,"深蓝"能搜索12步,所以它能战胜卡斯帕罗夫。"沃森"属于专家系统范畴(IF-THEN),通过对自然语言的处理和分析建立知识规则网,这与AlphaGo基于经验的推理具有本质的不同。

1.4 AlphaGo在认识论上的意义

值得注意的是,AlphaGo不仅在方法上实现了突破,而且在认识论上也有重要的进步意义,尽管还非常初步。主要有以下3点。

1)自学习和发现新知识的特点,展示了理解复杂事物的能力,具有帮助军事学、社会学、中医学等经验性学科建立科学体系的潜力。例如,通过搜集中医医学案例大数据,运用深度学习方法,可以挖掘、理解已有案例蕴含的经验,总结出中医的科学理论体系,有望使中医摆脱没有科学理论基础的困境。

2)AlphaGo能够发现人类尚未发现的着法,显示了部分领域在认知智能上机器也有超出人类的可能。过去一般认为计算机在认知智能上不可能超过人类,也不会具有棋感、直觉等人类独有的能力,但AlphaGo的出现,打破这种迷思。

3)人机对抗实现了"人""机"共同进步,而不是机器单向趋近人,揭示了未来智能化社会"人机智能"共生共进的前景。樊麾、李世乭与AlphaGo交手后都承认自己"长棋",能力得到提高;聂卫平从对AlphaGo不屑一顾到称之为"阿老师",也可以证明这一点。

AlphaGo虽然取得巨大进步,但它的机器智能仍属于工具层面的进步。只不过它可以是更聪明、更精准、更快速,甚至更全面的"工具"。因此,有人担心"AlphaGo证明计算机将统治人类",这个看法至少在目前看尚属多余,因为在可预见的未来,计算机还不会有"自主意识"。没有自主意识,就谈不上"谁统治谁"。

但是,从作战角度来看,AlphaGo却证明了未来"认知速度"必将成为战场关注的焦点。一般说来,过去的作战比拼的是机动速度、信息速度和火力速度,而未来战争在上述速度达到极限情况下,比拼的重点将转移到认知速度。一旦认知速度加快,就像是参加"快棋赛"一样,会彻底改变作战及决策思维模式。一些研究指出,一旦一方具有认知优势,对方可能只会像"傻子一样笨拙反抗",处处在对手的算计之中,失去灵活反应能力[22]。因此,从这一角度看,机器智能进步带来的将是一场"认知的革命",特别值得关注。

AlphaGo的技术可否用于兵棋?这是很多人关注的问题。有人认为,AlphaGo的方法很容易就可以平移到兵棋推演之中。但其实并没有这么简单。有两个问题,第一,兵棋是不是"棋"?是否能像"棋"那样对抗?第二,AlphaGo的技术可否用于兵棋系统?答案是肯定的。因为智能辅助对作战指挥、兵棋推演都极为重要。但哪些地方需要,又如何做到呢?以下将就这些问题展开讨论。

2 兵棋推演的瓶颈

2.1 兵棋推演的基本要素

兵棋是一种用于战争研究和训练的工具,主要分为手工兵棋和计算机兵棋。手工兵棋距今已有200年的历史,计算机兵棋始于20世纪60年代,80年代后期才随着计算机的普及被广泛应用[23-28]。

由于计算机具有超强记忆存储、综合计算、绘图表现等能力,可以更逼真地模拟实际战争,逐渐受到军方重视,并由此产生了两种军用计算机兵棋系统类型。一种是用于小组推演、想定作业等的小型兵棋系统,也常见于兵棋爱好者使用,但由于"一战一棋",因此很难具有普适性和均衡性。另一种是用于部队组织首长机关多方对抗演习的大型兵棋演习系统,具有兵棋实体多、行动种类多,演习内容全、系统规模大等主要特点,已经更多地应用于各国军队实际的军事演习。这两种兵棋系统目前都在使用,但大型兵棋演习系统更能体现兵棋研发的技术水平[27]。以下将主要以大型兵棋系统为背景展开讨论。

一般来说,一场大型的兵棋推演主要包括三大要素:一是参演人员,主要是参加兵棋推演的各级指挥员和指挥机关,例如各战区、方向、部队的各级指挥所;二是由兵棋系统模拟的战场环境和作战部队等,包括作战地域的地理、气象、水文、海洋、太空、网络等战场环境,各对抗方陆、海、空、天、网、电、核各级各类作战兵力和装备,以及不同方、各种部队和行动的各类作战数据、规则和约束等;三是负责组织兵棋推演的导演部及导调机构,负责组织兵棋推演,控制演习进程,讲评演习结果等。

这三类要素构成了一场兵棋推演的完整内容。兵棋的"棋盘"和"棋子"(图7),构成一个大的战场环境。例如,一场局部战争覆盖面积可以达到上千万km^2,在"棋盘"上就可能有近百万个六角格,每个格子还有约近百个属性。一场演习的兵力、保障、目标实体可能多达近10万个,实体种类可能超过几千种,实体属性多达上百个,而且可以分为十余方,所有的行动并行展开,与真实作战没有什么不同。从以上数据不难看出,这些要素的复杂程度很高,整个推演具有高度的动态性,而不是红蓝两方轮流出招的简单对抗。

事实上,兵棋对抗推演是按实际作战过程进行的,而不是按照"下棋"的方式"你一步我一步"进行的。一次成功的兵棋推演能非常逼真地反映实际作战过程,陆、海、空、天、电联合并行作战,各方根据态势进行不断地判断并做出决策,驱动整个兵棋推演持续推进,从而模拟出实际的作战流程及作战效果(图8)。由此可见,基于大型计算机兵棋系统的对抗推演需要模拟的作战要素繁多,关系错综复杂,对抗活动高度动态,是一个典型的复杂系统。

所以说,大型兵棋系统并不是"棋",因此将棋类对抗的方法用于兵棋推演,不仅显得过于简化不符合实际,而且目标也不一致。战争一个明显的特点就是"胜战不复",因为战争充满了大量的不确定性。也就是不会存在完全一样的两场战争。换句话说,同样的打法可能会是一场胜利而另一场失败。但棋类对弈不会这样,同样的过程一定会有同样的结果。因此,这就导致了兵棋推演的目标完全不同于棋类博弈,是"重过程而非结果,重发现问题而非胜负",这恰好是很多人不了解的。

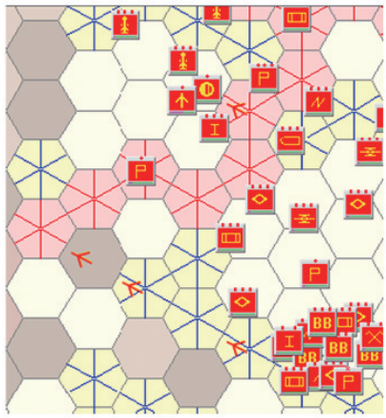

图7 兵棋的棋盘和棋子

Fig. 7 Board and pieces of wargame

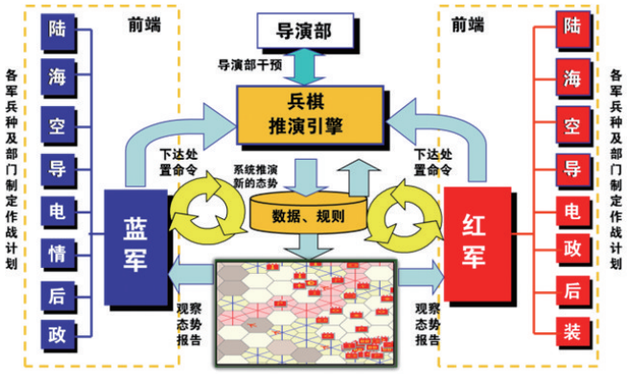

图8 兵棋推演复杂活动示意

Fig. 8 Complexity of wargaming

2.2 兵棋推演的难点在于模拟人的智能行为

兵棋推演的关键在于模拟"人"的行为,而这些行为都具有智能的特点。例如人对环境的判断、行动的选择、行为特点等,既包括个人行为,也包括群体行为。它不像围棋,围棋的棋子一落就不会再动了,而兵棋棋子则还会不断行动,去完成赋予它的使命任务。因此,兵棋推演主要面临以下5个方面的难题。

1)"假变真",即兵棋系统如何模拟对手。要让兵棋系统成为合格的"蓝军"或"绿军",能够按对手特点正确地反应、判断和决策。这种判断和决策可以简单区分为两个层次:一是初级部分,就是判断行动的意图,局部合理并且专业地应对,例如一发炮弹打来,该如何自动做出合理动作;二是高级部分,就是能按本方作战目标,自主地组织作战行动并力争取胜,一旦有此高级的全局决策能力,就可以真正扮演蓝方、绿方,甚至红方。如达到这个程度,本质上就和AlphaGo的目

标比较相近了,即创造出一个可进行人机对抗的“真正的对手”。此前类似的工作一直难以取得质的突破,根本原因就是受限于智能技术的发展水平。

2)“粗变细”,即兵棋系统如何模拟下级。将参演指挥员下达的战役命令,由系统转化为下一级可识别的战术、战斗指令,也就是要把上级的命令分解为下级各部队需要采取的行动。对于战略战役级的演习来说,上级的命令往往是概略的,因此就需要将宏观的“大命令”转化为一系列的“小命令”,来驱动下级指挥员和指挥机关的行动。“粗变细”还要将执行作战行动的高层级单位(如师、旅),转化为承担具体任务的小分队(如连、营、单架飞机、单艘艇船)的战术行动。这就需要兵棋系统具有能够理解上级意图,并转化为下级行动的智能水平。

3)“死变活”,即兵棋系统如何模拟兵力。主要是作战实体如何模拟出“拟人”的作战行动,而且这个行动的过程与部队的实际行动基本一致,具有智能化特点。例如飞机遇到防空火力会自动躲避,遇到雷场部队会自动停下来排雷;遇到不利于行动执行的气象条件会自动改变行动等。类似的这些功能虽然模拟仿真中也有,但由于判断条件过于简单,智能化水平与“人”相距甚远,演习中就有可能出现一些非常愚蠢的行动,主要原因就是机械地执行了事先设定的简单条件所致。战场环境是千变万化的,条件的衍生和组合又产生出新的条件,从而演化出天文数字的可能性,不可能由程序员一次就完全想清楚。

4)“静变动”,即兵棋系统如何模拟计划助手。美军有一个格言:“没有一个计划可以在接触敌人后还会继续存在”,说的就是仗一打起来所有的计划都会改变。因此,如何应对作战过程中层出不穷的不确定性,是作战任务规划、作战方案评估、兵棋推演等都面临的共同难题。战争是一个多方动态博弈的过程,一旦环境或者对手没有按照我方原先的设想行动,就必须要识别出这种变化,并重新决策或调整计划。以联合作战演习为例,通常需要事先制定包含陆、海、空、天等各军兵种行动的作战计划,但是,在实际作战开始后,只要在其中任一时间点上、任一军兵种的作战行动发生了变化,就可能会导致其后不可预知的无穷变化。因此,必须要求作战计划能随着战场态势的变化而临机改变。长期以来,虽然使用了指挥信息系统辅助制订作战计划,但仅靠计算机还无法识别各种临机变化,必须依赖人工来修改,这直接导致了“计划赶不上变化”。解决这个问题,就需要创造出能够识别态势变化,并实时进行计划调整的智能助手。

5)“无变有”,即兵棋系统如何模拟导演助理。大型兵棋推演过程和结果都异常复杂,即使是有经验的指挥员或专家,有时也很难一下就看清楚。因此,就带来难以量化评估,难以进行演习裁决等问题。这就需要创造出能够辅助导演进行导调的助理,其核心任务是协助导演判断各方态势、制权效果、作战目标达成效果等,从而为演习导调、评估、分析、裁决提供依据。

2.3 兵棋推演需要突破作战态势智能认知瓶颈

由此,兵棋推演5方面的难题,其核心可归结为两个问题:一是对战场态势的判断理解;二是对未来行动的正确决策处置。过去解决这些问题的主要方式,是起用大量有经验的“人”介入推演过程,弥补兵棋系统的智能不足。因而在兵棋推演过程中,会有很多参演人员来充当下级、蓝军、绿军等,更多的专业辅助人员协助转换、汇总、判断等。事实上,外军的兵棋推演同样也面临这些问题的困扰。

然而,就是全部用“人”介入替代,也不可能解决全部问题。一方面,并不是所有的“人”都是专业化的。信息化战争要求指挥岗位的专业性越来越强,蓝、绿军专业化不够就可能变成“红红对抗”,炮、装、工、通、化等指挥专业性不够就难以体现各兵种作战特点等。另一方面,很多情况下即使有“人”也难以正确处置。例如,作战结果的不确定性导致计划动态调整的不可预知,以及无人作战、Cyber作战等过程中无法回避的自主决策、快速行动[29]等,而且战役规模巨大,复杂性剧增,面对犬牙交错、相互叠加的战场态势,就是由“人”来看也不一定能看得清楚。这就像一个围棋高手面对复杂局面时,也不一定能算得清楚。正因为如此,这些问题已经困扰了人们几十年,仅靠现有技术一直得不到很好的解决。

AlphaGo展现出的认知智能上的进步,为众多方面应用展现了诱人前景。那么,AlphaGo的创新技术能否用于兵棋推演呢?答案很显然是肯定的。AlphaGo找到的并非是一种“下棋必胜”的方案,而是更精确的计算、更准确的判断以及更全面的分析,达到比“人类”更好的围棋如何取胜的理解。从这一点来看,它与兵棋推演是基本一致的。但战争过程更为复杂,充满了不确定性,甚至也不能简单用某种量化指标来衡量“胜负”。所以,兵棋推演“重过程而非结果、重发现问题而非胜负”的特点,要求兵棋系统的设计,必须首先解决前面提到的模拟“人”的行为问题,这是一个必须迈过的门槛。而其中,作战态势智能认知,则是难度最大的瓶颈性问题。可以这样说,解决了作战态势智能认知,就解决了兵棋推演中智能辅助的最大难题。

AlphaGo的突破已经展现了这个问题解决的可能,教会计算机看懂作战态势,判断形势并做出正确决策,也就是“感知理解”和“自主决策”,共同构成了作战态势“认知”最主要的环节。这一点与AlphaGo看懂棋盘、正确走子、争取取胜的模式非常类似。但不同之处在于,兵棋系统不仅需要得到正确结果,还特别强调正确模拟动态过程。事实上,在对抗推演情况下,取胜与否的结果要由对抗双方指挥员的指挥决策所决定。

“作战态势智能认知”技术具有颠覆性创新特点,是实现兵棋推演智能化突破的关键一环,也是兵棋系统升级换代的必要条件。这与一体化指挥平台中的决策辅助、作战规划系统中面临的智能辅助问题,在本质上是完全一致的。

3 作战态势智能认知——如何实现态势理解与自主决策

3.1 战场态势智能认知的难点

“作战态势”是指战争中对抗各方通过作战力量对比、部署和行动等形成的状态和趋势,指挥员通过态势进行指挥决策,作战结果和变化实时反馈又会形成新的态势[30-31]。战场态势有大有小:大的如战略战役态势,可以描述某个局部战争的战场时空状态,或者是某军种的态势,也可以是某个时节的态势;小的如某局部地域战斗态势,可以是只关注特定地域,也可以是只关注特定行动。

战场态势的层次不同,对态势认知的要求和内容也不同。但并不是态势越小,认知就越简单,越大则越复杂。而是各有各的问题,不同层次需要解决的问题和解决的方式都有所不同。此外,战场态势还有虚有实,虚拟态势如Cyber空间态势[32],实际态势如物理空间态势,两者在认知的难度上也有所不同。相比而言,虚拟态势增加了“人”认知的难度。

从作战指挥角度看,战场态势的理解过程通常可以划分为原始态势数据的收集、态势初级理解和态势高级理解3个阶段。首先,要获取各种信息、情报、报告等原始数据,对原始数据进行一定的整理,并呈现形成态势图、表格、文电等;其次,依据力量战损、空间布势、时间过程、环境约束等因素,进行态势初步理解,例如,估算部队的位置及关键部署、关键行动和关键时间节点等;最后,再进一步获得意图判定、优势结论、目标达成、效能评估、战局预测等更高层级的理解,例如,判断敌方的主攻方向、部队当前及后续的优势及劣势等。在此基础上,还可进行更高层次的进一步分析,最终形成对战场态势全面、准确、深刻的判断,并支撑指挥员做出相应的决策。

从AlphaGo的实现原理可以看出,AlphaGo做出决策前也需要进行态势的理解和判断。但是,要想直接套用AlphaGo的方法解决作战态势认知问题却是行不通的,主要原因是“围棋”与“战争”之间,在机器智能解决问题的条件上,存在以下巨大的差异。

1)信息条件方面的差异。完全信息条件还是不完全信息条件。围棋棋盘透明,是一种典型的完全信息动态博弈。围棋每走一步,都是根据当前的盘面信息作出的决策,并且当前的盘面信息对于棋手来说是完全知晓,没有任何隐藏。而战争过程并不是全透明的,甚至有时故意“隐真示假”,这是一种不完全信息条件下的动态博弈。也就是说,存在着战争“迷雾”,使得战争双方在决策时,无法把整个战场看清楚。显而易见,信息掌握越多的一方,越容易做出好的决策。另外,围棋的轮次规则可以保证在决策前态势不变,但作战态势则时时刻刻都在变化,决策只能在动态变化中完成,否则战机就会稍纵即逝。这种在“迷雾”中动态决策的要求,对人来说都很困难,对人工智能来说更是难上加难。这种差异,必然会使得数据存在大量噪音,甚至有很多数据是假数据,想用这种数据训练出像AlphaGo一样的深度策略网络,显然是非常困难的。

2)规则确定性方面的差异。确定性规则还是不确定性规则。围棋规则是确定的、清晰的,规则对于双方也都是相同的,双方的起点实力也是相同的。而在战争领域,由于对抗双方实力可能不对等、对抗因素不确定,使得对抗规则也不对等。例如,伊拉克战争中,美军可以超视距攻击伊拉克的防空阵地,也可以用钻地弹攻击位于地下掩体的目标,而伊拉克则不行。同理,恐怖组织可以无视国际法袭击平民目标,而正规军队则一般不会采用这样的行动。这些都是作战规则不确定的体现。所以,这种战争规则不对等、不固定、动态变化等特性,相较于围棋的确定性对等规则来说,更难以把握和运用。

3)训练样本数量上的差异。机器学习需要大量的样本,但围棋和战争两者的样本数量具有天壤之别。围棋具有大量样本,AlphaGo要得到16万盘人类对弈的数据并不困难,而且数据采集相对容易,只要将每一步双方的行动方案记录下来即可。但战争数据来源少,即使是历史上已有的战争数据,也不具备所有的属性,都是简化的记录,数据质量不高。而且,通过各种演习、试验积累的数据,也还是近几年的事情。这一点上,美军和以色列军队做得比较好,实战的数据比较多,但数据的来源众多,大多数是异构数据,应付机器学习的需要仍很难。因此,要想在机器学习上有多突破,必须另辟蹊径。

4)计算量方面的差异。计算量主要取决于博弈空间的范围和决策选择的内容。围棋的计算量已经巨大,需要上千个CPU、GPU才能应付。但战场空间的计算量更要远远超出围棋的量级范围,如果仅在大规模计算能力上下功夫,显然也是很难解决问题的。事实上,在实际作战中,指挥员处理复杂的战争问题主要依赖于专业训练和指挥经验,而专业训练又依赖于军事理论知识和战争经验的长期积累。因此,找到快速缩小博弈空间的方法本身,就是智能算法首先要解决的问题,以应对兵棋推演时爆炸式增长的计算量问题。

3.2 解决作战态势智能认知问题的可能途径

要让计算机理解战场态势,基本思路就是基于深度学习的态势认知,即利用历次作战、演习的数据,进行逐层训练得到深度神经元网络,逐步得到对复杂作战态势的理解,并在此基础上进行准确判断和合理处置,实现辅助或自主决策。但需要强调的是,作战态势极具复杂性,态势理解是一个复杂的系统工程,仅使用深度学习方法还难以解决所有问题,还要借助更多智能技术手段才能完成。有一些方法已经展现了较好的前景,还有一些实际案例也值得关注。

1)增强学习。AlphaGo之所以能战胜人类,其最重要的创新就是采用了两种不同的技术方法训练策略网络:一种是有监督学习的策略网络,它仿照人类的套路走法,从大量棋谱数据中学会下棋,属于经验招式;在此基础上,又通过自我对弈、不断进化,得到了更高级、更优化的策略网络版本,即“增强学习策略网络”。

增强学习实际上是一种自我强化的训练方法(图9)[33]。首先由机器做出一个行动,再由外界给机器一个反馈,机器通过这个反馈来评价自己行动的优劣,最终反过来调整优化行动的策略[33-35]。通过不断反馈和调整,针对某一个态势,机器能利用学习到的经验,给出最优化的行动。因此,增强学习本质上是一种与人类获取经验非常类似的方法。AlphaGo正利用增强学习技术,得到了具有更强能力的版本,并训练出了估值网络[2]。此外,DeepMind公司依靠深度增强学习技术进行像素级的训练,也已经在47种视频类游戏中超过了人类玩家[36-37]。

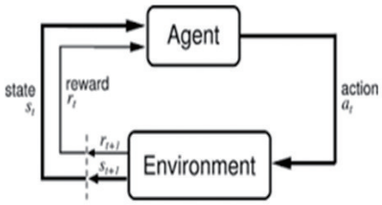

图9　增强学习原理示意

Fig. 9　Principles of reinforcement learning

2)样本生成。自我强化训练需要大量的样本,但作战态势数据样本太少,就需要找到一种利用少量样本产生出更多样本的方法。AlphaGo训练估值网络时,为解决样本不足的问题,就采用自我博弈产生了3000万盘新的棋谱数据。正是通过创造出的大量随机样本,AlphaGo才有了"超出人类已有样本的基础";利用大量随机样本进行自我博弈训练,才做到了"超出人类的训练"。由此,它才有可能做到具有"超出人类所有经验的能力",战胜了人类。

因此,生成出大量样本,也是深度学习在样本稀缺时的必然选择。其他研究领域在解决此类问题时也采取了类似的做法,比如图像旋转以生成更多的图像学习样本等[38-39]。解决作战态势样本不足问题,也可以采用类似的方法。例如,可以采用局部空间、时段分解、动态随机推演等方法生成新的样本。兵棋推演以实战化作战想定为背景,通过兵棋模型来逼真地模拟作战环境和作战过程。因此,可以将兵棋推演看成是一种作战样本数据的生成方式,用得到的仿真数据作为样本数据的重要来源。运用兵棋推演生成数据还有一个好处,就是可以根据研究需求设计想定条件,这也是一种样本按需生成的方式。所以,样本生成从本质上讲,是一种"以少代多"的样本获取方式。

3)迁移学习。迁移学习(transfer learning)是在具有相似性的领域,利用已经训练好的模型得到新模型的一种学习方式[40]。它只需要再增加一小部分新领域的数据,就可以很快完成训练并用于新的领域。这种学习方式有点类似于人类所具有的"触类旁通""举一反三"能力。例如,在教会机器看懂"坦克"之后,再教它认知"自行火炮",它的学习过程就会更快、更容易,不用再像认知"坦克"那样需要非常大量的数据。一旦能够在相似的领域实现迁移学习,不仅可以做到相似场景模型的迁移,还能够做到不同形式数据之间的知识迁移。例如,先让计算机学习"听"电影的语音,再让计算机学习"看"电影的时候,计算机识别电影视频中图像的能力就会增强。这种方式对于战场态势的认知非常重要。

4)小样本学习。人类对事物的认知,具有"只看一眼就可以识别"的能力,即可以通过"例子"抓住特征,实现小样本学习。例如,人只见过一次菠萝,下次就可以很快在众多水果中挑选出菠萝,这就是抓住了最重要的特征。但机器可能需要学习几十万张菠萝的图片才能正确识别,花费的时间太多。

针对这一问题,Lake等[41]提出了一种名为"概率规划归纳计算结构"的方法,采用基于案例的贝叶斯[42]方法,通过小样本学习来模仿人类笔迹(图10[41]),并通过了图灵测试。因此,这种小样本学习是一种"照猫画虎"的认知方式,具有典型的人类概念抽象的特点,对作战态势的理解也有重要的借鉴意义。

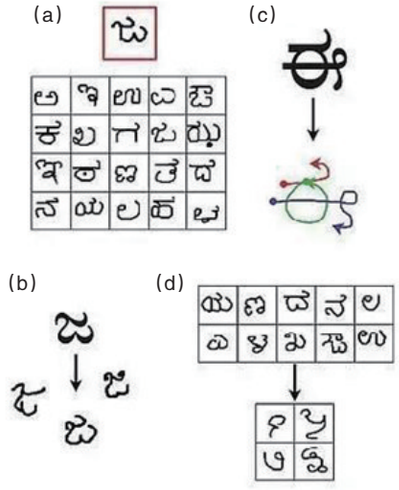

图10　计算机模仿人类的笔迹

Fig. 10　Computers imitate human handwriting

5)"星际争霸II"人机对抗。"星际争霸II"是2010年出品的即时战略游戏,可多人对战,在特定的地图上采集资源、生产兵力,并摧毁对手的所有建筑而取得胜利。该游戏本质上也是在不确定条件、不完全信息条件下,进行多兵种行动和策略的对抗,因而实现这一人机对抗的核心基础,也是对战场态势的认知,只不过游戏态势是对真实战场态势的简化。

此前有报道称,DeepMind公司正与开发"星际争霸"游戏的暴雪公司协商,宣称将挑战"星际争霸II"人类高手。如果挑战成功,意味着作战对抗智能认知技术将会更进一步,它所构建的作战态势神经元网络的认知理解可能更加完善,并且很容易迁移到真实的作战态势认知方面,距离作战态势智能认知的突破可能只有一步之遥,对此绝对不能忽视。

6)"德州扑克"人机博弈。DeepMind团队还在做德州扑克的人工智能[43]。德州扑克属于不完美信息动态博弈,对抗各方的牌面不完全公开,要求对抗各方在部分信息条件下进

行决策和对抗,这增加了计算机理解和决策的困难。他们采用深度增强学习(DQN)技术和虚拟自我博弈(NFSP)[44]相结合的技术,在没有任何先验知识的情况下,训练出了2个包含4个隐含层的神经网络。这2个神经网络的作用与AlphaGo的策略网络和估值网络非常相似,工作机制也与AlphaGo相似。这种方法运用到德州扑克人机博弈时,神经网络虚拟自我博弈能够非常好地逼近纳什均衡[45],目前已经达到了人类高手的水平。

德州扑克与AlphaGo虽然都训练了类似的神经网络,但两者的训练方式和使用方式却略有不同,AlphaGo策略网络的训练使用了人类的棋谱经验,而德州扑克没有使用人类经验,没有任何先验知识,是完全依赖于增强学习自我博弈获得的经验,也就是说,德州扑克完全是"自学成才"。另外一个不同,是德州扑克没有使用蒙特卡洛树搜索算法,而是将策略网络和估值网络得到的策略进行了混合策略求解,类似于博弈论中混合策略纳什均衡的求解方法,德州扑克的这种搜索方法,使其策略选择的速度要比AlphaGo快得多。

7) 空战格斗人机对抗。近期,美国辛辛那提大学与空军研究实验室合作开发了一个叫"Alpha AI"的机器飞行员[46]。在模拟对抗演习中,Alpha AI击落了所有其他机器飞行员,并在与美国空军战术专家基纳·李上校的人机空战格斗对抗中大获全胜,展现出了巨大的优势。Alpha AI的核心技术是遗传模糊树[47],它是一种结合了遗传算法和模糊控制的智能控制新技术,在处理大规模复杂的智能系统问题方面具有卓越的性能。Alpha AI在空中格斗中具有更快的基于态势智能感知的战术计划速度,可以比人类飞行员快约250倍(图11)[46]。此外,Alpha AI还能够以集群的方式控制大批空军无人机,在格斗中快速收集敌机的信息。即使在模拟过程中,研究人员故意限制Alpha AI所配置的武器系统能力,使其处于劣势,它仍然能够最终击败人类飞行员。这在战术层面展现了智能态势认知带来的速度上的优势,值得高度关注。Alpha AI的出现意味着人工智能将会很快走入实战领域。

图11 Alpha AI的快速战术计划界面

Fig. 11 Tactic displaying Alpha AI utilization of flanking the opponent

3.3 战场态势智能认知研究的基本思路

人对态势认知的过程,是一个整合信息并形成决策的过程。同样,利用机器智能实现态势认知,也应该分层次、分阶段地按照相应的步骤去实现各自的功能。为了降低战场态势理解的难度,可以按照时间特征、空间特征,分别将其分解成多个层次来进行处理,这也是一种便于快速突破的简化、聚焦的思路。例如,按空间层次,可以将态势理解的内容分解为位置关系、军种关系、指挥关系、保障关系等特征;按时间层次,可以将态势理解的内容分解为不同时节的力量消耗、编配部署、任务完成等特征。

战场态势认知的核心是理解"整体关系"。虽然分层理解可以简化认知过程,但最终需要整合成一个对整体关系的理解和整体态势的把握。在将分解后的层次态势整合为包含整体关系的整体态势过程中,由于战场态势固有的复杂系统特点,需要注意以下关键问题,一是局部与整体不具线性可加性,即分解后的层次态势与整体态势之间并不是线性的累加关系,局部理解是基础,但整体态势并不能仅由局部态势简单叠加而来;二是态势演化过程具有非线性,对抗效果的判定,需要考虑态势的变化率,而非仅凭一时一隅的得失;三是态势认知需关注全局性,包括对关键节点、关键任务、关键效果的判定,都需从全局角度出发,而不是仅从某一支部队、某一次行动的角度判定。上述这3个问题,本质上都是从局部低层次的态势,涌现出全局性的高层态势过程中,无法回避的复杂性问题。

在对战场全局态势的正确理解基础上,要解决的下一个智能认知问题就是行动的决策,即行动策略的智能化选择。从算法的角度来看,行动策略的选择一般可划分成为广度搜索和深度搜索两个过程,AlphaGo围棋落子方案的选择也是由这两个过程构成(图12)。其中,广度搜索重点解决胜负直觉的问题,深度搜索实际上是做一个推演估算,就是推算后续几步会出现什么情况。这就像下棋,人类总是首先通过直觉对盘面的每个点进行估计,选择有利的几个点位,即进行宽度上

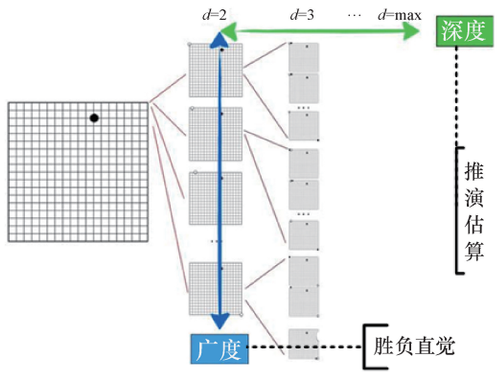

图12 决策选择的深度搜索和广度搜索

Fig. 12 Deep search and breadth search in decision-making

的搜索,然后在有利的点位上进行深度的推演估算,最终确定哪一个点位最优落子。因此,无论是兵棋系统、任务规划系统还是方案评估系统,决策选择都需要从深度和广度两个角度来加以计算。但需要强调的是,战场态势理解和作战行动决策是并行进行的,并不是一个简单的串联关系。

由于态势认知的高度复杂性和特殊性,解决作战态势智能认知问题,需要“多管齐下”,多种方法混合使用。通过深入分析可知,包括AlphaGo在内很多成功的机器智能应用产品,都是众多人工智能技术的组合。AlphaGo就是蒙特卡洛树搜索技术、深度学习技术、增强学习技术的融合。具体来说,在学习算法设计方面,具有大量数据且特征模糊时,可用深度学习和增强学习方法;当只有少量数据但特征清晰时,可以考虑采用基于案例的贝叶斯方法。在决策选择方面,可以结合AlphaGo的蒙特卡洛树搜索、“沃森”的自然语言知识网、“深蓝”基于规则的暴力搜索等多种方法。在数据来源方面,可以从历史数据、仿真演习、靶场试验、生成器生成等途径中获得不同认知研究需求的数据。此外,由于态势认知涉及陆、海、空、天、网电等各个作战空间,可以在不同的作战空间分别开展智能态势认知,如美军从2010年就开始发展电磁空间中的“认知电子战”技术,目前已经取得巨大进展,极大地提高了美军电磁空间威胁的认知响应能力。

3.4 作战态势智能认知的研究意义和实现载体

作战态势智能认知技术是兵棋系统研发中急需突破的关键技术,是因为它具有以下重要意义:1)它是一个具有颠覆性、基础性特点的科学问题,一旦突破会带动一片问题的解决;2)它是可以与战斗力直接挂钩的实用性课题,通过“认知速度”提升作战指挥效能,能够显著提高战斗力水平;3)它是作战相关信息系统升级换代的核心课题,一旦解决可以促进兵棋演习系统,也包括指挥信息系统、作战规划系统等的全面升级;4)它是对手必然封锁、我方必须自主的课题,任何人都不会提供他所得到的作战神经元网络;5)这不是地方公司重点关注和研究的课题,必须要由军方自己完成。美国防部2007年6月提出的“战争算法”的概念,目标是利用人工智能技术从更多的信息源中获取并处理大量信息,为作战提供数据响应建议,搭建从数据到决策的桥梁,这在本质上就是智能态势认知技术。

现在已有的研究基础包括以下4个方面:1)AlphaGo等已经证明了技术的可行性;2)在复杂网络、大数据、深度学习研究方面已经有了一定的积累;3)多年信息化建设和兵棋演习积累了大量数据;4)建立了一批军事、技术复合型的人才队伍。特别是军队改革提出了迫切的需求,以前部队对兵棋推演提出的改进要求,大多都属于需要智能化突破的要求。

可以用兵棋系统中的“智能蓝军”和“智能参谋”作为研究载体,实现突破。所谓“智能蓝军”,就是让系统充当“蓝军”“绿军”参加演习,将会变得更加专业,可以在一定程度上实现自动反应,从而实现局部的人机对抗。所谓“智能参谋”,就是实现系统充当部分下级部队或部门并且不被察觉,完成识别态势、分解任务,调整计划、自主决策、监控行动等功能。如果能够做到这一点,就可以认为兵棋推演的智能辅助达到了新的水平,在作战态势智能认知方面有了突破。

4 结论——人机智能是兵棋推演的新机遇

智能辅助是制约兵棋系统升级换代的瓶颈问题,是一个不容忽视甚至是需要争分夺秒去解决的问题,因为这关系到作战的需要。但这个问题的难点就在于“态势认知”,即让计算机正确理解作战态势并做出决策。也就是说,需要得到能够理解作战态势并作出应对的神经元网络。考虑到作战态势的复杂性,“人机结合”可能是更适合的方式,也就是将人的优势和机器优势相融合来解决问题,而不是让计算机全部取代人,称这种方式为“人机智能”方式,它在2个方面与以往的理念有所不同。

1)计算机可以不再是单纯地辅助人决策,而在有些情况下需要替代人决策。这是因为,有了机器智能的辅助,计算机可以不仅仅是“辅助人更好地决策”,或“得到更好的作战方案”,而在某些方面可能比“人”做得更好,这时就需要把决策权赋予它。比如在高速的Cyber作战行动中,人根本无法跟上行动过程,这时就必须让计算机替代人去决策。这就像AlphaGo与李世乭的比赛,决定权并不在那个负责投子的黄世杰手上,AlphaGo提的也不是建议,而是最后的决定。所以,在这里机器智能的目的不仅是“帮人”,而是要“超人”。

2)计算机智能与人是可以共同进步的,也即机器智能会促进人类智能的进步。这是AlphaGo给人们的重要启示,其结果就是人与机器的结合使得双方都进步。因此不能停留在计算机模拟人类智能的阶段,人们也应向计算机智能学习,机器智能会进一步促进人类智能。反过来说,计算机做一些事情也远远没有人做得好,所以如何做好人机分工,也非常重要。这样,未来作战通过人机智能辅助,不仅能够满足“人”的要求,也能得到更好的作战结果,而这个结果是人与机器合作得到的。这是一个巨大的改变,谁先意识并做到这一点,谁就将在未来战争中占据主动。

未来的战争胜利将取决于认知速度,而“认知速度”的快慢,取决于对智能技术的掌握程度。绝不要低估机器智能技术突破门槛后的发展速度,2015年还曾有人认为,“在围棋上计算机战胜人类尚需100年”,但仅仅几个月后AlphaGo就超过了人们的想像。智能技术的特点是,突破它的门槛很高,可能会很长时间都被挡在门外。但一旦突破,它反而会有借力加速的作用,使得发展普及的速度很快。如果哪个国家在作战态势认知上取得突破,抢先跨入门槛,就会导致其他国家在战略战术上都非常被动。

同时,也不能急功近利,关键是搞好基础性研究。AlphaGo团队坚持了10多年,终于在围棋上战胜了人类。中国通过在兵棋系统研发上的努力,必将在智能辅助瓶颈问题上取得突破。

参考文献(References)

[1] Campbell M, Jr A J H, Hsu F H. Deep Blue[J]. Artificial Intelligence, 2002, 134(1/2): 57-83.

[2] Silver D, Huang A. Mastering the game of go with deep neural networks and tree search[J]. Nature, 2016(529): 484-489.

[3] Allis L V. Searching for solutions in games and artifcial intelligence[D]. Maastricht Netherlands: University Limburg, 1994.

[4] Newnan M E J. The structure and function of complex networks[J]. Siam Review, 2006, 45(2): 167-256.

[5] Wang F Y, Zhang J J, Zheng X H. Where does AlphaGo go: From church turing thesis to AlphaGo thesis and beyond[J]. IEEE/CAAJournal of Automatica Sinica, 2016, 3(2): 113-120.

[6] Hopfield J J. Neural networks and physical systems with emergent collective computational abilities[J]. Proceedings of the National Academy of Sciences. 1982, 79(8): 2554-2558.

[7] Andrade M A, Chacón P, Merelo J J, et al. Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network[J]. Protein Engineering, 1993, 6(4): 383-390.

[8] Anguita D, Gomes B A. Mixing floating and fixed-point formats for neural network learning on neuroprocessors[J]. Microprocessing & Microprogramming, 1996, 41(10): 757-769.

[9] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.

[10] Schölkopf B, Platt J, Hofmann T. Greedy layer-wise training of deep networks[C]//International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2006: 153-160.

[11] Arel I, Rose D C, Karnowski T P. Deep machine learning: A new frontier in artificial intelligence research[J]. IEEE Computational Intelligence Magazine, 2010, 5(4): 13-18.

[12] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521 (7553): 436-444.

[13] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//International Conference on Neural Information Processing Systems. South Lake Tahoe, Nevada, USA: Curran Associates Inc, 2012: 1097-1105.

[14] Jaderberg M, Simonyan K, Vedaldi A, et al. Reading Text in the Wild with Convolutional Neural Networks[J]. International Journal of Computer Vision, 2016, 116(1): 1-20.

[15] Clark C, Storkey A. Teaching deep convolutional neural networks to play go[J]. Eprint Arxiv, 2015: 1766-1774.

[16] Tesauro G, Galperin G R. On-line policy improvement using monte-carlo search[J]. Advances in Neural Information Processing Systems, 1996(9): 1068-1074.

[17] Browne C B, Powley E, Whitehouse D, et al. A survey of monte carlo tree search methods[J]. IEEE Transactions on Computational Intelligence & Ai in Games, 2012, 4(1): 1-43.

[18] 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 perspectives in military application of artificial intelligence[J]. Journal of Command and Control, 2016, 2(2): 114-120.

[19] Sutton R, Barto A. Reinforcement learning: An introduction[M]. Massachusetts: MIT Press, 1998.

[20] Kimura H, Miyazaki K, Kobayashi S. Reinforcement learning in POMDPs with function approximation[C]//Fourteenth International Conference on Machine Learning. Sydney: Morgan Kaufmann Publishers Inc. 1997: 152-160.

[21] Hasselt H V, Guez A, Silver D. Deep reinforcement learning with double Q-learning[C]//Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix, Arizona USA: American Institute of Aeronautics and Astronautics, 2016.

[22] Alberts D S. The agility advantage: A survival guide for complex enterprises and endeavors[M]. Washington: CCRP, 2011: 23-71.

[23] Peter P P. The art of wargaming[M]. Annapolis Maryland: Naval Institute Press, 1990: 1-21.

[24] Peter P P. Ed Mcgrady. Why wargaming works[J]. Rhode Island: Naval War College Review, 2011, 64(3): 111-133.

[25] Blank Dennis. Military wargaming: A commercial battlefield[J]. Jane's Defence Weekly, 2004(2): 5-9.

[26] Rose D. Designing a system on system wargame[R]. Ohio: US Air Force Research Laboratory, 2006.

[27] 胡晓峰, 司光亚, 吴琳, 等. 战争模拟原理与系统[M]. 北京: 国防大学出版社, 2009.
Hu Xiaofeng, Si Guangya, Wu Lin, et al. War gaming & simulation principle and system[M]. Beijing: National Defense University Press, 2009.

[28] Caffrey J, Mattew B. Intelligent computing and wargaming[C]//The International Society for Optical Engineering Orlando, Florida: The Society of Photo-Optical Instrumentation Engineers, 2011: 5-9.

[29] Musman S, Temin A. Evaluating the impact of cyber attacks on missions[J]. M&S Journal, 2013(7): 25-36.

[30] Endsley M. Toward a theory of situation awareness in dynamic systems [J]. Human Factors, 1995, 37(1): 35-64.

[31] Oosthuizen R, Pretorius L. System dynamics modelling of situation awareness[C]//Military Communications and Information Systems Conference. Piscataway, NJ: IEEE, 2015: 1-6.

[32] Tadda G, Salerno J J. Realizing situation awareness within a cyber environment[J]. Proceedings of Spie, 2006, Doi: 10.1117/12.665763.

[33] Sutton R S, Barto A G. Reinforcement learning: An introduction[J]. IEEE Transactions on Neural Networks, 2005, 16(1): 285-286.

[34] Kimura H, Miyazaki K, Kobayashi S. Reinforcement learning in POMDPs with function approximation[C]//Fourteenth International Conference on Machine Learning. Netherlands: Morgan Kaufmann Publishers Inc. 1997: 152-160.

[35] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.

[36] Bellemare M G, Veness J, Investigating contingency awareness using Atari 2600 games[C]//AAAI Conference on Artificial Intelligence. Washington: AAAI, 2013: 864-871.

[37] Mnih V, Kavukcuoglu K, Silver D, et al. Playing atari with deep reinforcement learning[J]. Computer Science, 2013, arXiv: 1312.5602.

[38] Ji S, Xu W, Yang M, et al. 3D convolutional neural networks for human action Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 221-231.

[39] Jayanth Koushik. Understanding convolutional neural networks[J]. Computer Science, 2016, arXiv: 1605.09081v1.

[40] Pan S, Yang Q. A survey on transfer learning[J]. Knowledge and Data Engineering, IEEE Transactions on, 2010, 22(10): 1345-1359.

[41] Lake B M, Salakhutdinov R, Tenenbaum J B. Human-level concept learning through probabilistic program induction[J]. Science, 2015, 350 (6266): 1332-1338.

[42] Assael J A M, Wang Z, Shahriari B, et al. Heteroscedastic treed bayesian optimisation[J]. Computer Science, 2014, arXiv: 1410.7172.
[43] Heinrich J, Silver D. Deep reinforcement learning from self-play in imperfect-information games[J]. Computer Science, 2016, arXiv: 1603.01121v1.
[44] Iii T J L, Epelman M A, Smith R L. A fictitious play approach to large-scale optimization[J]. Operations Research, 2003, 53(3): 477-489.
[45] Ponsen M, De Jong S, Lanctot M. Computing approximate Nash equilibria and robust best-responses using sampling[J]. Journal of Artificial Intelligence Research, 2011, 42(1): 575-605.
[46] Ernest N, Carroll D, Schumacher C, et al. Genetic fuzzy based artificial intelligence for unmanned combat aerial[J]. Journal of Defense Management, 2016, 6(1): 1-7.
[47] Cordon O. A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems[J]. International Journal of Approximate Reasoning, 2011(52): 894-913.

AlphaGo´s breakthrough and challenges of wargaming

HU Xiaofeng[1], HE Xiaoyuan[1], TAO Jiuyang[1,2]

1. Department of Information Operation & Command Training, National Defense University, Beijing 100091, China
2. College of Command Information Systems, Army Engineering University, Nanjing 210007, China

Abstract This paper summarizes the principles, new methods, technological breakthrough, and the epistemological sense of AlphaGo. Then the bottleneck of intelligent wargaming is analyzed, and the significance of intelligent situation awareness in wargaming is addressed. Next, the way to realize situation awareness in operations is proposed. Finally, new challenges of man-machine intelligence for wargaming are discussed.

Keywords AlphaGo; deep learning; wargaming; situation awareness

（责任编辑　傅雪）