

学术笔谈

编

今年既是“人工智能夏季研讨会”召开60周年,也是认知科学诞生60周年。前不久,谷歌公司的一款围棋人工智能程序AlphaGo以4:1战胜了韩国九段棋手李世石,更将人们对人工智能的关注推前到前所未有的高度。一时间,全世界都在热议“机器战胜了人”“人类智力的最后堡垒坍塌了”。事实果真如此吗?人类被机器人取代、人工心灵全方位超越人类心灵是必然趋势还是危言耸听?如何看待人工智能的发展现状和未来可能?各种热评逐渐退潮后,我们需要冷静下来,认真思考与人工智能相关的问题。本版今天特刊发3篇文章,希望能够引发读者的思考与讨论。



如何理解人工智能

□ 刘晓力



算的问题,人们无论如何找不到算法。而且不可计算的世界相比可计算的世界要大得不可胜数。

国际计算理论、视觉理论家马尔曾可将计算之事归结为三大步骤。第一,要把具体的现实问题抽象化成一个可以清晰定义的问题;第二,要看它是否是可计算类的问题;第三,要在可计算的范围内找到算法。对于人工智能而言,自然还有重要一步,就是要在机器上执行算法完成计算。目前人工智能可以在很多需要大量计算的局部领域超出人的计算能力,但它的局限性最尖锐地体现为不能处理以下三类问题:意义理解;意识和意识体验;自主性和知觉。

第一个问题可以归结为,机器理解语言吗?我们知道,人类是借助抽象符号使用语言来描述外部世界的,人是世界意义的制造者,这个意义就是我们内在的心理世界的所思所想与外部世界的关系。机器懂得人类语言表达的意义吗?机器翻译程序真的懂得所翻译的语言内容吗?手机上安装的语音输入系统真的理解人们所输入的语言符号的意义吗?这一直是人工智能的所谓“瓶颈问题”的困难所在。

时至今日,机器根本不能理解语言的意义,不能理解我们跟外部世界究竟有怎样的关联,因为这个意义的解释者只有人类自身。当前人工智能最重要的进展,一方面是机器在感知外部环境信息方面能力的提高,如iCub、谷歌大狗、谷歌自动驾驶汽车和飞机等;另一方面是通过多层神经网络深度学习,机器在模式识别、图像识别、语音识别能力上有令人惊喜的表现。但是可以说,目前的各种识别还只是完成某种“匹配”任务的过程中并不关涉图形和语音所包含的意义。

第二个问题是,在与世界打交道的过程中,人除了有感官知觉,还有情感、意识和意识体验。人们把李世石输棋的原因归为他的情绪不稳定,但这恰好就是人类超过机器的独特之处。人有喜怒哀乐、爱恨情愁;有追求真理的欲望;有宗教情感和同情之心,能在与他人的交流中获得理解、体验痛苦和快乐;人也经常

会出错会失败,有各种非理性的行为。这些正是人之为人的价值所在。机器没有这样的信念欲望意图和意识体验。一个当下处理围棋问题的确定算法一经运行就不会更改,所以机器不会出错。如果有错,责任也在编写算法的人。AlphaGo既没有恐惧也没有焦虑,没有喜也没有悲,它是靠着强力搜索高速计算取胜的,在这一点上,恰好说明作为算法的AlphaGo有不及人类智能的弱势之处。尽管目前人工智能专家也在研究机器情感、机器意识,但是离机器真正有情感有意识和意识体验,恐怕还有本质上的距离。

第三个问题是,单纯作为算法的AlphaGo没有自主性和知觉。AlphaGo的运行需要外部驱动力,这个动力是人给予的,围棋算法是人设计的,实现算法的程序是人编写的,3000万个棋局是历史上棋手长期积累又由人陆续输入数据库的。此外,更重要的是,人的行动是有动机的,人有意图会规划未来,有做出选择和行动的自由意志,更别提棋感、直觉、欺骗、故意下坏棋了,它的行为只受输赢这个简单的目标导向。从比赛结果看,像是这样一个硅基材料的机器有了胜过世界冠军下棋的智慧和能力,但事实上,依照我们下面的分析,AlphaGo的智能实际上是以人为核心的一个巨大的延展认知系统的能力,是系统部分认知功能的某种外化。

AlphaGo 不仅是一个算法,还是一个延展认知系统

如前所述,如果仅仅把AlphaGo看作一个算法,它显然是无法与人类的心智相匹敌的。那么,我们究竟在什么意义上可以解释它的智能呢?这个解释的角度就是两位世界级哲学家克拉克和查尔默斯1998年提出的延展认知和本体论、认识论

通常,人在环境中生存,依赖环境判断采取行动。人要在环境中即时处理特定问题时,自然会选择周边所能供给的环境条件、物理设备、已有的知识背景和环境网络为我所用。例如医生在诊断病情

最初,当人们预测AlphaGo和李世石围棋大战谁会赢时,我就断言AlphaGo会赢。因为李世石面对的不仅仅是一个算法程序,除了包括走棋网络、估值网络、快速走子程序和蒙特卡罗搜索几大部分构成的复杂算法之外,人机大战的战场还装备了40个在线网络平台、1200多个CPU、170多个GPU,还有大数据、云计算即时处理当前棋局和史上积累的3000万个棋局,甚至还有一条专门铺设的赛事光缆,而这一切都是一群世界一流的深度学习专家所为。从这个意义上讲,AlphaGo是一个人机交互的巨无霸耦合体,是一个即时处理专项任务的延展认知系统,这个系统的核心是人。前表演的只是这个系统的程序运行终端,而人才是其背后的智力资源。

无心的机器是目前人工智能的局限

AlphaGo战胜李世石产生的一个直接效应是,很多人认为,AlphaGo的胜利意味着人工智能已经突破某种极限达到了高智能水平,照此下去,人类所珍视的所有智慧终将被机器所替代,人类未来命运堪忧。

但事实上,虽经历一个甲子岁月的洗礼,人工智能目前还没有摆脱“无心的机器”的命运,今天的人工智能可以说既无智力也无心灵。这是因为目前计算机使用的都是二进制编码的图灵机计算,是递归算法。80年前,伟大的逻辑学家哥德尔就已指出图灵机的计算能力与人类相比是有先天局限的。但是,人们更热衷于引用图灵1950年在《机器与智能》中的话,“我们不能因为一台机器不能参加选美大赛而责备它,就像我们不能因为一个人没有飞机飞得快责备他一样,机器也能够思维”;却忽视了1936年图灵经典文献《论可计算数》的基本点,其中清晰地区分了什么是可计算的、什么是不可计算的,以及计算的界线在哪里。只要是不可计

人工智能的春天真的来了吗

□ 危辉

除了AlphaGo以4:1战胜了李世石,近年来,类似的例子还有不少。例如,IBM公司的Watson程序在问答比赛中战胜了人类冠军,计算机写新闻稿和赋诗、自动驾驶、所谓意念控制无人机等。由此似乎可以得出结论,人工智能的时代真的开始了,世界将面临被机器接管的风险。

我们在接受这些突破之前有必要探究一下隐藏在背后的技术实质。例如,Watson是基于文本匹配的检索,没有多少推理的成分;赋诗程序是效仿标注格式的充填,需要人工筛选;自动驾驶所依赖的附加设备价值超过车辆本身好几倍,且对环境敏感;所谓意念只不过是分离出来的脑电信号,对它的分类结果在可靠性、稳定性和可重复性上都很差。

AlphaGo下棋程序在人工智能领域属于计算机博弈这一研究分支,那么,它究竟难不难实现呢?双人博弈下棋游戏开展得非常早,属于人工智能少数几个鼻祖级的研究对象之一。例如,上海科技馆有一个下五子棋的机器;1997年IBM公司开发出了国际象棋程序;2008年的西洋跳棋程序更是达到了人不犯错就只能与它下平手的水平。可见,人工智能学者在这一领域已经深耕了很多年。

所有这些双人博弈游戏都有三个共同特点:第一,游戏规则非常明确;第二,棋局布局很清晰且很完整;第三,棋盘空间很有限。这些看似不重要的特点对计算机程序、对人工智能来说却是极端重要的利好。因为正是这些规则性、规范性和有限性,使得看似复杂的下棋问题在计算机上变得非常可行,如此明确的计算对象恰是计算机编程能够如鱼得水的理想环境。所以,计算机下棋虽然不是挑战,反而远比自然语言理解、场景理解、非线性推理等任务要容易研究得多,否则它就不会在人工智能诞生的一开始(20世纪50年代)就成为我们的研究对象之一。

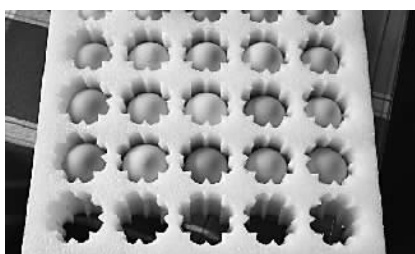
在技术实现上,人工智能下棋程序通常使用一种被称之为“状态空间搜索”的方法,其核心思想是:第一,把棋局的演化过

程看成一个个状态,用某种数学形式进行记录;第二,由下棋规则规定的走步方法可以把这些状态串联起来,形成一张非常庞大的状态演化网络;第三,用搜索的办法在网络中找出对自己有利的走步策略。这种方法由于形式化程度高,因此非常适合在计算机上编程搜索。当然,这也带来了存储空间和搜索效率的问题。实际上,我们往往会利用一些有价值的信息来引导搜索的方向,如我们“歧路寻羊”时会利用羊留在路上的蹄印或留在草上的牙印来筛选跟踪路径。这次AlphaGo下围棋同样会面临这样的问题。它利用机器学习技术,从大量现成的对弈中学到了两样至关重要的东西,一是如何由当前的棋局映射出若干应子对策,二是如何对对应子后的布局进行利己性打分。前者使它能够快速找到对应走步,后者使它能够快速对不同走步的好坏进行判断。这样就避免了它在庞大超乎想象的围棋状态空间(理论上总共有3361个状态)中进行穷举搜索,只需进行小范围的有限搜索即可,其中所谓机器学习过程就是从大量以往的对弈棋谱中建立当前布局与下一步走步之间的对应关系。AlphaGo所基于的搜索技术和机器学习技术其实都是现成的,但它在对棋局模式的刻画方式和如何对棋局好坏进行评价方面还是有创新性的。尤其是它对围棋布局的数字化表征方式,这直接影响到机器学习效率和能否最终找出布局与走步之间的对应关系。

AlphaGo的成功更像是人工智能领域中基于实例的推理研究,也就是针对当前面临的问题去寻找以往经历过的一个或几个类似的问题,把当时的解决方案拿来做一个适应性修改后应用到新问题上去。当套用到下围棋这个需求上时,这个基于实例的推理过程就可以简化为当前棋局布局与以往某个布局间的相似性匹配和走步方法类推。因此,我们可以说谷歌公司用一种非常恰当的技术解决了一个非常有显示度的问题。但

是据此说这个围棋程序能够“学习”和“创新”还是值得商榷的,因为它实现的实质上是一同范畴下的类比。它进行走步决策时完全不理解人类棋手那些“金角银边肚皮”之类的下棋原则,也不需要懂什么叫“实地”,只是把下围棋当作了一个由一种布局模式推演到另一种布局模式的过程,只不过其带来了存储空间和搜索效率的问题。我们甚至不能认为这是一种智力,充其量是一种强悍的细分记忆力。

人类围棋下输了,那么人类智能就此崩塌了吗?其实人类智能的强项可能不在下围棋这种只有极少数人才能达到高水平的项目上,而在那些应付不可预期的能力上面。例如下图,它显示了一种包装鸡蛋的方式。



假设我是第一次碰到这种方式的包装,且要从这样的盒子里把鸡蛋拿出来,我的第一方案是把鸡蛋抠起来,但我的背景知识马上告诉我这样做不行,因为鸡蛋会被抠破的。我的第二方案是把整个包装的一边移出桌子的边缘,然后从上往下捅,但细想这也不保险,因为稍不注意就捅到地上去了。由此我立刻想到我们可以从下往上顶鸡蛋,取鸡蛋的问题就顺利解决了。这样的问题看起来不难,几乎人人都能解决,但其奥妙的地方就在于类似这样的问题我们可能是第一次碰到,要解决它时并没有预编程,也就是说我们脑子里没有针对性地储存在好如何拿鸡蛋的程序,而是靠灵机决断做到的。大量的背景知识帮了我们大忙,知识引导我们如何去行动。这样的例子有很多,如我们所熟知的乌鸦喝水、曹冲称象、司马光砸缸等,都堪称求解问题的典

时会借助医疗仪器,科学家在实验室工作时会利用实验设备,一个不擅长记忆的人出行时需要手机、笔记本电脑或网络信息的引导。克拉克认为,这时,这个行动者就与这些外部条件构成一个动态的延展认知系统,那些即时可利用的设备就可以看作延展认知系统的一部分。实际上相当于在处理特定的问题时,把个体的智能的一部分功能卸载到外部设备上,这个外部设备只要能跟他即时地连接,一起参与完成当下认知任务,就实现了某些认知功能和心智能力。但是,到底选用什么样的外部环境设备,以及如何主动与外部世界互动,显然有一个主动性的问题。处在环境中,人的感知器官和运动系统,以及整个有机体与世界构成一个关联体,成为一个耦合的认知者和行动主体,其行动是整个系统的动态行动。如果考虑的是在社会认知情景中的系统,问题将更加复杂。

在今天这个信息化技术高度发达的时代,人的很多认知功能,认知负担都已卸载到物理设备上,而且还在逐步卸载到各种便携式、穿戴式和嵌入式“人类增强技术”的技术产品上,许多人类的感知功能和体验甚至可以在虚拟现实载体上延展地实现。因此,在笔者看来,说作为机器的AlphaGo打败人类,就相当于说与人耦合的这些外部设备超越人类智慧一样。也就是说,在以人的认知为核心的,可以展开大规模并行计算的,包括CPU、GPU、云端大数据,甚至包括光缆和人类下围棋的历史经验整合在一起的延展认知系统在有限时间战胜了围棋冠军李世石,这还是比较冷静而客观的说法。这种系统的智能是一种集群式智能,是多个个体和多台机器网络、大数据平台和云计算平台即时连接的并行集群智能,它是有因果历史链条的,算法只是它的最近一端。至于这样的智能系统的认知能力和智能的特征及其本质,它与人类智能的区别何在,包括由此所产生的延展认知系统的本体论、认识论问题、延展认知系统的社会伦理问题,以及延展认知系统与人类未来命运的关系问题等,恰好是即将到来的人机交互的新时代必须面对的哲学课题。

(作者单位:中国人民大学哲学院)

那么,人工智能的现状是什么?我们以前经常讲,人工智能的发展遇到了瓶颈,笔者认为,“瓶颈说”对于人工智能的现状而言还是很高的评价,因为起这码说明我们已还把瓶子的别处都看了个通透,并且能够确定何处才是出路所在了。但事实上,人工智能还达不到这个程度。笔者认为,人工智能的现状是“瞎子摸象”,每个人的观点和看法都不同,缺乏整体性的认识,因此也就没有长远的解决方案。人工智能界经常引用这样一段话——“飞机能够成功上天是因为我们放弃了对鸟的模拟”。既然如此,人工智能为什么非要效仿人类智能机制呢?这句话还引出了一个隐藏在现实之中的弊端,那就是,人工智能界现在已经分不清什么是权宜之计,什么是长久之策了。

(作者单位:复旦大学计算机科学技术学院认知算法模型实验室)



人工智能

性特征和接受性特征。而这种心灵的主观性方面本质上是与一种独特的视角即第一人称视角相结合的,并且也是仅仅在心理活动组成的;知觉是我们通过感官通道获取外界信息的活动;行动是我们心灵输出的部分;意识是伴随着这些心灵模块的工作存在的,通过意识我们可以知晓并说出自己当下的心理状态;自我意识是一种将自我作为对象的认识。对于心灵,我们可以理解为包含这些模块的整体。而当前人工智能的主要工作,涉及的是认知层面。当然,现在感知智能的研究与产业化也日益兴盛,如语音和图像处理得到了迅猛的发展;至于行动智能,近几年也有一些机构做得非常好,如Alphabet公司的波士顿动力。但实际上这些做法很大程度上还是基于认知的人工智能来实现的。最近,又有研究通过模拟人类的大脑这一更为基础性的方式来“全方位模拟人类心灵”。加拿大滑铁卢大学做了一个名为Spaun的机器,它里面有250万虚拟神经元,是目前世界上最大的功能性大脑模型。Spaun可以做简单的工作,比如模仿人类大脑进行看、听、记忆等基本认知活动,也可以做一些拷贝行为,比如模拟作画,还可以犯类似人类的错误,用类似人类的的速度处理一些问题。最近,该研究的主持人以利史密斯甚至认为人工心灵的时代就要到来。虽然相较于人类千亿美元数量的神经元,现在机器的神经元还很少,但是它所具有的基本能力已经表明了它未来企及甚至超越人类心灵的可能性。随着技术的进步,谷歌、IBM、高通等一些企业的研究的深入,深度学习与控制理论等一系列理论的进展,以及来自各个政府、企业的一些资助项目的推动,包括欧盟的人类大脑计划、奥巴马政府的大脑计划以及在世界各国政府在推动的各项相关计划,以利史密斯相信,应该可以做出全方位模拟人类大脑的机器。他还给出了未来50年人工心灵的进展时间表。这个时间表显示了在冯·诺依曼架构和神经模拟架构上的人工心灵在每个阶段(比如5年、10年、15年内)可以实时处理的神经元数量和可以执行的心灵功能:在刚开始的5年里,这样的机器所做的工作很有限。比如,它对所识别的视觉和听觉信息的质量要求比较高,它只具有非常基本的语言理解能力,它的行动也很受限。随着时间的推移,其所用的神经模块越来越多,它的能力也会越来越强。25年后,冯·诺依曼架构在物理学层面受到一些限制,神经模拟架构的做法更具优势。这时候机器已经可以表现得基本类似于人的心灵。50年后,人工心灵就可以全方位超越人类的心灵。然而,人工智能究竟能否像以利史密斯这样的乐观者预测的那样,会企及人类心灵并超越人类心灵呢?对此,笔者将从人类心灵的本质这一维度来考察机器究竟是否具有人类心灵,人工智能和人类心灵之间有没有什么界限或鸿沟。我认为基于人类心灵和人工智能的本质不同,这两者之间可能存在如下七条鸿沟:主观性与客观性、第一人称视角与第三人称视角、私密性与共享性、个体性与集体性、离线性与联机性、情绪性与无情性、可错性与无错性。主观性是人类自身无法摆脱的一种心灵特征。思想、知觉和行动必然是归属于某个主体的,包含了主体的自发

吗

那么以上这些鸿沟产生的根源又是什么呢?换句话说,人工智能和人类心灵最根本的差异在哪里呢?人工智能可以模拟人类心灵的很多重要特征,比如自主性。但这仅仅是模仿,因为自主性概念本质上还是排斥模仿的。而且,自主性还会涉及我们的行动和自我意识,而自我意识对于人工智能来说又很难实现,因为数据的完全共享性与自我意识的个体性难以兼容。另外,我们的认知、知觉和行动模块里面又存在现象性。即使我们了解清楚了大脑的一切构造,在我们的物理世界里面也无法做出这种现象性的东西来,这就是所谓的意识的“难问题”。自主性和现象性可以说是人类心灵所独有的特征,是人类心灵与人工智能之间更深层次的差异,它根植于人类个体的肉身性和数字性的可共享性之间的根本不同,而这正是两种心灵的异质性的基底。正是这种肉身的存在使人类心灵具有了自发自为的自主性来维持这种存在,也正是这种肉身的运作使人类心灵具有了独特的现象性。

总之,从目前人工智能的进展来看,我们很大程度上做到了让机器具有认知之心。当然也有些没有做到,如认知里面的现象性。自主之心、现象之心、意识之心,这些都是目前人工智能很难处理的问题。由于人类心灵和人工智能的根本性的不同,要让机器具有类似于人类的心灵也几乎是不可可能的。(作者单位:南京大学哲学院)