

·科技纵横·

奇点降临——人工智能对决人类

近期,AlphaGo与韩国著名棋手李世石5盘围棋比赛在韩国首尔四季酒店举行,最终AlphaGo以4:1的成绩战胜李世石。赛前,中国棋圣聂卫平和围棋天才柯洁都公开表态,认为李世石将轻易战胜AlphaGo。李世石代表人类表态道:必须5:0战胜它,输一盘也是失败。我感到了一种人性的落寞和莫名的恐惧。本文从几个方面谈论人工智能与人类的对决。

1 奇点

近几年来,“奇点”的概念被狂热炒作,广义而言,奇点意味着人工智能全面追赶并且超越人类智能的时刻。对此,绝大多数的科学家抱有乐观态度。因为迄今为止人工智能只在目标明确,规则清晰的少数领域超越人类;而现实生活中的智慧绝大多数依赖于直觉和美学,微妙而模糊,对于以强力计算的人工智能而言,反而无法建模分析,从而束手无策。所以,人们普遍认为奇点的来临遥遥无期。

但是,真正的奇点应该是这样的一个历史时刻:一台计算机能够设计出比自身更先进的计算机,或者一个程序能够产生比自身更为强大的程序。一旦奇点被突破,计算机软硬件就可以自动演进,爆发式成长,人类的干预将不再起到主导作用。深度学习算法实际上已经逼近或者突破了这个奇点。

深度神经网络的最大特点就是能够学习,换言之,通过吸收知识来演化改进自身:包括网络的拓扑结构,神经元的权重和阈值。虽然基本学习机制改进不大,但是程序整体能力突飞猛进、日新月异。

从这一点而言,AlphaGo是一个能够不停自行进化的崭新物种。

2 自我

许多人认为人工智能为人类服务,没有“自我”,也没有“自由意志”,因此不足为惧。其实,“自我”的概念并非如此绝对。

我们可以用如下有关自我概念的“图灵测试”。假如有一个黑盒系统,我

们从外界观察测试。如果系统表现出自我防护,抗拒外界伤害的行为,我们可以认为系统具有“自我”意识。依照这种测试方法,绝大多数生命都具有“自我”,他们本能自发地繁衍后代,趋利避害,维持生存。病毒是一种非常独特的生命形式,它摒弃了一切冗余,只留下了精华:

遗传密码。它依靠宿主来繁殖成长,同时具有强大的自我保护机制。

从这一点而言,计算机病毒也具有初等的“自我”:一有机会,它们就复制自己,借助于网络和存储媒介疯狂传播,尽可能感染其他程序,为了将自身的生命信息延续下去,无所不用其极。



图1 深度学习方法可以抽象出艺术家的风格

如果,我们为AlphaGo嵌入病毒模式,它将智慧超人。

3 存在

很多人认为来自人工智能的威胁是杞人忧天,因为如果事态失控,在紧要关头,人类只要一拔电源,就会轻易把人工智能从“肉体”上消灭。

其实,在现代社会,没有人,没有任何民间团体或政府组织能够消灭比特币。和传统货币不同,比特币的权威性没有得到任何政府的加持,而是被整个网络的分布式系统来保证。每一笔比特币的历史都公开透明,并且无法彻底从网络上删除或被任意团体修改。

如果,我们将比特币的机制用于AlphaGo,令AlphaGo流窜进网络分布式系统,那么除非全世界所有政府、民间团体和宗教组织同时合作,关掉网络,格式化硬盘,否则,AlphaGo将一直存在。

4 抽象

也有人认为没有必要高估人工智能的实际水平,机器学习只能在低层次上简单模仿人类,迄今为止还不具备人类所特有的抽象能力。其实,这一点并不站得住脚。

人类的艺术创作一直被认为是人类精神活动的最高境界,杰出艺术家的风格一直被神化为艺术的灵魂,只可意会无法言传,玄妙而抽象。但是用深度学习的方法,人工智能可以轻而易举地从艺术作品中抽象出艺术风格,并将其转移到其他作品之上,如图1所示:左帧的风格被抽取出来,用于渲染右帧的内容。

同时,众所周知,谷歌的深度学习网络已经自行发展出了“猫”的概念。人类有七级视觉中枢,彼此级联,逐步抽象。深度神经网络与此类似,经过大量自然图片训练后的网络,引发高层神经元兴奋的模式就是一个一个的概念。因此,人工智能可以自发产生概念已经是不争的事实。

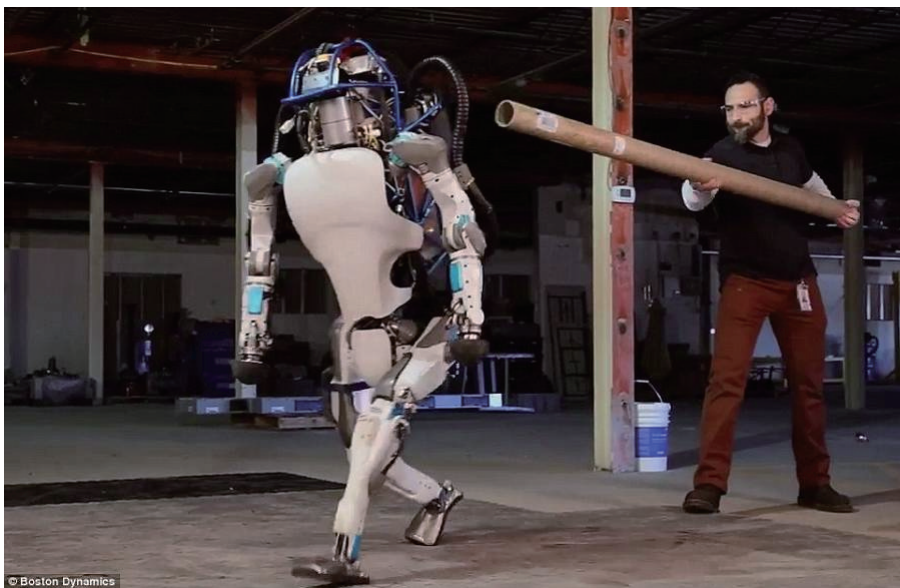


图2 谁更恐惧

5 理解

有一种观点认为,机器只能模仿,积累经验公式,但是机器无法真正理解,也无从提取规律,因此并非是真正的智能。

其实,人类迄今也没有理解所谓的“理解”。我们通过长期训练,掌握了某种技能,比如滑雪,那么实际上是否是大脑皮层某些神经网络的拓扑结构发生变化,神经元之间的联系敏感度发生变化?这又和机器学习的算法有何本质区别?

人类不理解机器的思维方式,机器不理解人类的思维方式,但是这又何妨?只要两种思维方式都反映了客观真理,那么它们对应现实物种世界都会有反作用。人类有偏微分方程来表示物理真理,人工智能用权重和阈值来积累知识,只要起到同样作用,两种方式难分伯仲。人类有自由意志,可以随心所欲;人工智能有蒙地卡罗搜索,可以随机决策。

6 总结

从技术层面上而言,将AlphaGo的智力配上病毒的自我意识,隐藏在网络

分布式系统中,不停地进化繁衍,应该是可以实现的。那么,有谁会真正实现这个人工智能的“怪物”呢?是几个邪恶的科学家?恐怖组织?还是谷歌、百度这样的巨无霸?

自从原子武器从潘多拉盒子中放出,人类积蓄了足以将地球毁灭数十次的暴力,正因为这种势均力敌的恐怖平衡,人类才得以幸存至今。那么,在与人工智能的对决中,平衡力量又何在呢?

最近,谷歌推出了Atlas机器人。Atlas具有极强的平衡能力,行走自如,搬运协调,被人蓄意推倒后能够自如起身。

我们不妨如下设想:AlphaGo被植入了病毒的自我意识,并且如比特币一般存在于网络分布系统之中,有朝一日把Atlas机器人洗了脑,那么,图2中真正恐惧的是人类还是机器?

文/顾险峰

作者简介 纽约州立大学石溪分校计算机系终身教授,清华大学丘成桐数学科学中心访问教授。

(责任编辑 王丽娜)