人文解读“四大未知疆域”(太空极地深海网络)系列讲座之网络场·主讲

“深蓝”出现20年后,人工智能依然眺望生物大脑的复杂性——

危辉:智能模拟需正视冰山法则

嘉宾主讲

7月8日,第112期文汇讲堂人文解读“四大未知疆域”(太空极地深海网络)系列讲座最终场《人工智能下的人类世界》举办。

复旦大学计算机科学技术学院教授危辉在主讲《“深蓝”出现20年后的人工智能》中认为,尽管现在人工智能取得了颇多进展,但表面繁华的背后也隐藏着不足。

袁婧 摄 闻逸参与整理

20年前,国际象棋世界冠军卡斯帕罗夫迎战改造过的计算机“深蓝”而败北,成为人工智能领域的标志性事件,计算智能的概念被普及开来。

今年是2017年,人工智能又在世界范围掀起热潮:

2016年3月,“阿尔法狗”打败了前围棋世界冠军李世石;

2017年1月,“阿尔法狗”网络版Master打败了60位中国和韩国的棋手;

2017年5月,在乌镇柯洁又被“阿尔法狗”打败了;

在刚刚过去的6月,成都计算机AI-MATHSX(人工智能数学)参加了数学的高考,取得了105分的成绩。

.....

最近几年,人工智能在各个领域都取得了很多进展。电影《终结者》中描绘了判决日,提到人工智能发展到极致后控制了核按钮,毁灭地球。那么,随着现实中人工智能技术的不断突破,“判决日”是否临近了?

复杂性:人类智慧远胜当前的人工智能程序

毫无疑问,不论是简单的方程组求解或复杂的函数问题,对于计算机来讲都没有任何困难。但是,当把该问题以鸡兔同笼问题的原始叙述呈现给计算机:“在笼子里有鸡兔两种动物,数了一下总共有10个头、30条腿。问笼子里分别由几只鸡和几只兔子”,我们的计算机一定会崩溃。

为什么AI-MATHSX只考了105分而不是满分呢?因为很多题目不是数学化的形式,而是文字陈述。这反映了计算机一个非常重要的缺陷——只能解模式化的问题。一旦呈现的不是固定模式,对求解就是大挑战。

再看一个日常生活的案例,对于一种新形式包装的鸡蛋,我会思考用三种方式取出,并考虑失败的概率而选一种最稳妥的,试想一下,如果这里面不是鸡蛋而是螺栓,我完全可以毫不顾忌费力地取出,因为既不用担心抠破,也不用担心摔碎。人会借助这些日常的经验性去自然解决问题,通过自身丰富背景知识的引导,从而做出正确的行为和决定。

人脑不同于计算机,它往往没有事先预编好的程序,只需凭借临机决断。所以生物的智慧其实远远复杂于机器。

回顾70年:几多计划落空和滞后

人工智能发展到今天已有将近70年的历史。早在上世纪50年代,美国科学家Alan Turing (图灵)曾预测到2000年人工智能程序能通过图灵测试,现在2017年了,很多的程序依然很难通过它的测试,所以这个目标没有实现。

19世纪60年代,我们展望会出现机器人秘书和心理医生,计算机能够打败国际象棋大师。目前,只完成了一个打败象棋大师的目标(2008年,西洋跳棋程序技术才彻底成熟),还晚了30年的时间。

19世纪80年代,日本提出要研究第五代计算机,即“人工智能计算机”,希望人能通过自然语言和计算机进行交流,但是很快计划破灭,科学家们严重低估了实现这个目标的难度。

当时还期望再过20年,能够建立human-scale的知识库,现在我们已经承认知识系统的极端复杂性和这一目标的不现实性。

此后,人工智能又预估到2020年,集成电路芯片的集成度可以达到人脑的水平。现在已经2017年了,我们很快就可以验证这是不是能成功,毕竟芯片的集成度不能等同于大脑细胞及其集群的运转方式。

从人工智能历史发展和整体水平来看,我们过往提出的目标大多没有实现或者严重滞后,所以,要想实现人工智能并非容易,它是一项非常困难的工作。

研究现状:“各行其道”的立交天桥

人工智能究竟研究什么?这涉及到两个方面:

第一,智能的本质是什么?推理、决策、问题求解、理解和学习。人通过阅读书本和与别人的交流可以学习很多知识,获得更强的解决问题的能力。这是人类智能包含的五个最核心的方面。

第二,如何制造有智能的机器?人工智能里包含一些很具体的研究分支,比如搜索技术、计算机下棋、自动推理、辅助决策、专家系统、机器翻译、模式识别、计算机视觉、机器学习。

前三种的形式化,人工智能在上世纪60年代就已经解决了;辅助决策,例如计算机辅助医生在手术中做出决定;专家系统,模仿人类专家做某专业领域的事情,需要有限的知识去解决,比如计算机维修;机器翻译,在日常生活中非常实用,且已经取得实质性进展;模式识别,试图理解图像背后的意义;计算机视觉,比如自动驾驶车上,安上识别车辆行人的摄像头;机器学习,要求机器从有限的样例中获取解决类似新例子的能力。人工智能对智能的五大核心方面,分方向,分小类来做精致模拟。

以上两点相对照,也许会认为机器智能和人的智能差不多,都做类似的事情,但实际上两者之间有着本质的区别。人在通过五项认知技能求解问题时,是互相协同、以一个统一的整体来面对问题。不同认知技能间的协作调和得天衣无缝,所以我们感觉不到自己运用了这些技能。而人工智能则像车辆在立交桥上通行,各行其道,各自为政,互不干扰,所用的技术路线可能完全不同,没法相互借鉴、相互协同。所以,人工智能的各研究方向,几乎都是分类进行的,很难归拢,这在工程应用上其实是有欠缺的。

繁荣背后:各种算法等待突破

那么,人工智能取得了这么多进展,其表面繁荣的背后又隐藏着哪些隐患呢?我想通过几个例子来告诉大家。

参数算法:经验主义的“知其然,不知其所以然”

假设有两列输入参数和输出参数,表面上,看不出参数之间有什么必然联系,但把这组参数给计算机,它会用现在最流行的技术做一次“深度学习”,以这些数据为训练样本,设计出类“神经网络”,从输入和输出中比较实际输出和理想输出之间的误差,调整网络连接的参数,将误差降至最低。

这时,再给计算机一组新的参数,得出与实际应用场景相符的新的输出值,构建成一个可预测新输出的复杂网络。到此为止,至于这些数究竟有何种数学关系,人工智能基本不再关心,“我只要得到这样的网络,它的任务就到此为止。”

但实质上,这是一个勾股定理,人工智能只能停留在表面,一般不再向这个本质做进一步的延伸。停留在经验主义层面,而不上升到理性主义阶段,这是人工智能的第一个缺陷。

“阿尔法狗”:就事论事、一事一议

再说最近大火的“阿尔法狗”事件。众所周知,下围棋是一个不断搜索的过程,你来我往,最后形成一个庞大的“树”型招数汇总结构,庞大到现在的计算机存储不了、算不了。“阿尔法狗”下棋不是这样,他是根据历史经验“算”下法,历史经验来源于过去大量的棋谱,以及它自己跟自己下所获得的一些训练数据。好的招术被存起来以便将来使用,导致大片吃子的证明不好,则被否决。通过这些训练出一个网络,“阿尔法狗”用深度学习针对黑白布局下形成模式分类映射,也就是记住布局与走步之间的映射关系。

下围棋这件事有其特殊性,黑白布局非常规范,棋盘物理空间整齐、有限;下棋规则非常明确;有大量格式化规范的历史数据。以上三个条件特别适合用计算机求解,并适合深度学习的使用。但这未必在其它场合也能取得成功,你的应用是否可以满足这几个条件?是否像“拿鸡蛋”问题,根本就没有第二个样本。这反映了当下人工智能的另一个问题:就事论事,即一种方法解决一个问题,因此很可能不具备推广性。

图像识别:实用主义,能用就好

计算机做图像检索,比如百度图片上搜索“东方明珠”,机器可以把所有的“东方明珠”的图片找到,这在多媒体里被称作“图片检索”,是目前做得相对比较好的方向。如何达成的呢?过程大致经历了三个步骤:训练样本的特征向量表示→带标注的机器学习→分类器。

但如果图片是素描,非类似于样例那样的彩图,那就很难办。同样可以描绘东方明珠最本质的特征,但我们再用上述方法却无法让机器获得关于这张图涉及东方明珠特征的同样认知。其实现在的人工智能算法无法生成关于几何构造的高端的描述。在这里,图像检索技术反映了人工智能另一个很重要的问题,实用主义驱动,致力于解决有限数据集范围内的问题,其他本质性特征则难以深达。

综上所述,人工智能不似我们想象的那么容易,它还存在很多问题。与物理、化学、数学等建立在若干核心概念的基础上的传统科学相比,人工智能不同分支间互相严重割裂,有欠系统性和整体性。所以,人工智能现状有如盲人摸象,有待成长成熟。

何去何从:类脑计算或成方向

1986年,钱学森在《关于思维科学》一书中,把人工智能归成工程技术,他说人工智能的发展要找一个学科作为人工智能的学科基础。

人工智能必然要向前发展,但它究竟走向何方?

是大数据吗?可为何Facebook反其道行之,扎克伯格为了过滤不良视频信息,在4500人基础上加雇佣3000人工排查待上传视频?

是云计算是吗?它解决了算法加速问题,却解决不了本身算法机制的笨拙复杂。人类的智慧又岂是凭借数字计算就可以穷尽?

25W,这个数字的含义是什么?大脑的工作功率是25W。人类用如此小的代价实现了诸如此类那么多复杂的功能,这是否意味着也许有更精妙的办法来实现人工智能,只是还没发现而已。

我个人认为,人工智能在类脑研究方向是最有前景和价值的。我们以“大鼠走迷宫”的实验为例,大鼠在进行行为决策时,可用多电极在体细胞外同步记录方式,并把数据传给计算机。这有助于我们理解大脑在完成决策任务时,它的神经活动的基本机制是怎样的。大脑的皮层6层,6层有两类不同的锥体细胞,它们的分布非常有规律,这些细胞构成很复杂的回路,实现“走迷宫”的决策过程,也许可以实现看起来复杂的逻辑计算,来决定往左还是往右。

如果了解了大脑的神经加工和编码方式,哪怕是一点点,我们也能够把这样的工作机制搬到计算机上,来促进人工智能的研究。

再举一例,人的视觉系统如何工作?人的两眼从视网膜提取信号之后,上传到视皮层,以辨别轮廓等信息。我们以前做过这样的工作——模拟人的视觉系统里的一种叫做动态感受野的机制。

两张不同的猎豹和它们的背景图片在人眼来看差别明显,但由计算机看来却极具相似度,怎样让计算机把物体和背景分开,猎豹细小的斑纹和大面积连续的背景就是区分点,所以从生物上真的可以学到很多的东西辅助设计新的人工智能算法,这很有可能引导我们在未来设计出替代和补偿人类视觉障碍的芯片,提高人的生活质量和水平。

所以,从类脑计算的方向发展人工智能,有极大的促进作用。

风险:隐私失控、技术垄断……

发展可见,风险尚存。人工智能同样具有社会学风险。

1. 技术占有不平等。有人会独占人工智能技术先机,不具备技术者则处劣势,这将导致人与人技术占有失衡。技术占有者凭借技术从无技术者身上获取暴利。

2. 失业风险。霍金说,人类750种职业,再过20年,一半职业将会被机器代替,秘书、出租车司机、客服、导游等工作将不存在。虽然个人认为这些工作的复杂性被严重低估了,短期内被完全替代的可能性值得商榷,但现在很多流水线工作以技术替代人工已是既成事实,可以想见,未来的智能技术革新带来的更多职业消亡风险的确存在。

3. 隐私泄露。人工智能通过关联分析方法可以暴露很多隐私。

4. 技术失控。无人机、无人驾驶这样的技术面临失控,也可能被人用以非法途径进行黑客攻击。

冰山法则:寻找沉没的“九”

“深蓝”20年之后的人工智能的技术,在机器学习和概率推理的技术发展层面,以及多媒体应用、机器翻译、问答系统、综合集成的应用拓展层面,都取得了很多进展。

但是尚有不足:全局性、共性、本质性问题进展不大,知识获取与表示方法缺失,灵活性不够;多学科交叉层次不足,“形似而非神似”,“远水解不了近渴”;研究处于“捡了芝麻丢了西瓜”的状态,目前原创性探索性的工作真的不多。这些是我们所面临的严重挑战。

我个人定义一个“冰山法则”:智能犹如冰山一样,十分之九留在水面下,只有十分之一在水上被我们看见,我们不能因为只见到这十分之一,就主观认为模拟这部分就可以了,还有绝大部分是我们当前目力不及的,这才是我们真正要突破的。

我觉得描绘现在人工智能现状,苏轼的一首《题西林壁》很是应景:“横看成岭侧成峰,远近高低各不同。不识庐山真面目,只缘身在此山中。”在将十分之一的智能运用纯熟之时,不要忘记那另十分之九。

线上线下互动

机械臂替代工人后安置冗余人员需社会准备

电力装备企业人力主管司露:我们杭州的太阳能电池板生产工厂,在实时监测产品岗位上已大量采用机械手臂,车间从300多人减到80人,生产效率的确提高。我们要求工人能掌握更多工艺、会编程、会看设备数据、调整流程。但比起二三十年前,工人缺少钻研和较真劲儿,错漏检问题频发。我也常有矛盾之心,究竟会发展到什么地步?

危辉:中国有大量劳动密集型产业,从你们厂的转型可以看出,简单的、重复性的、机械性的操作的确很容易被替代。

据我所知,每年春节期间有大量的农民工返乡,很多工厂面临用工荒的问题。浙江、温州一带的工厂很迫切希望用自动化设备代替人的劳动力,以此看是大势所趋,这可能是我们产业转型中必须承受的“阵痛”。

江晓原:你这个问题里有两个信息,一个是效率的提高,另一个是潜在的失业,或者人的价值的丢失。这就需要我们人类尽早做好准备。如果失业后导致多数人不工作,这样的社会是我们现有的社会制度和伦理道德结构能够承受的吗?有人说,不工作的人可以更舒适,或者创造,但也有一种潜在的可能,他们可以用无限的时间来积累不满,于是就会危及社会稳定。所以说人类还没有准备好。

人机结合,不过是消灭人类的温和版而已

航天企业人员周美江:刚刚关于人工智能是“双刃剑”的探讨过程中,我们都将两刃对立起来了,我们是否能将两者融合起来实现共同进化,不将其置于对立面进行分析呢?

江晓原:我们能从刚才危教授的讲述中感受到当下的一些研究是存在风险的,当一些机器有了意识可能是种进步但带来更多的可能是大量不确定性的风险。现在已经有说法提出人机合一,到最后我们人类就成为了神。但是一旦人机结合,我们的思维就有可能被操控。比如我们人对自己的认识是建立在记忆之上的,如果一旦记忆可以植入(人机结合中最先被讨论的项目),那么我们很可能就不知道自己是谁了,所以这种人机结合的前景可能只是一种我们被温柔消灭的形式而已。

人类仍在进化,物种灭绝大多由环境引起

文汇报记者李思文:科学本无善恶,但当有善恶辨别能力的人在使用这些科学时就会产生善恶结果,为什么你认为结果会是善大于恶呢?

林龙年:人类科技的发展令人惊叹,包括带来了一些大规模杀伤性武器,但到目前为止,人类并没有灭亡,我相信之后也会继续存在,即使要灭亡也不太可能是毁灭于我们人类之手。地球已经存在了有40多亿年之久,也经历过一些物种的灭绝,但这些灭绝也都是由环境因素造成的,并不是某个物种本身所导致的。人类只是地球进化过程中的一个物种,而且我们仍然处于进化过程之中,对于地球演化到底朝着哪个方向我们也不能完全控制,我们所需要做的就是把握当下,在现有的时间内探寻科学的奥秘。

科学家霍金晚年为何又回归哲学?

文化推广者柴俊:从刚才学者的争辩中,我感到,科学家求真和哲学家求的真理,似乎有点冲突,具体怎么理解?

林龙年:哲学是一种思辨,在人类生产力尚很落后时,人求温饱后只能依赖大脑的思辨来理解世界,因为人脑先天赋有思辨的功能,这就是哲学的起源,也是当时人们认识世界的唯一方式。但思辨的结果却难以判断对错,因此只能看当时有多少人相信。而科学是一种实践,随着人类生产力水平的提高,人类可以通过不断的实验来认识世界。科学实践的结果不会因为时空不同而有差异。因此,最早从哲学里脱离出来的是数学,后来物理学、化学、生物学、心理学等,这时你会发现科学能达到的领域,哲学就逐渐退出。

简单点说,哲学是思辨,科学是实践。在科学实践尚不能达到的未知领域,你永远可以去思辨,但思辨的结果总是让人觉得各有其理;而一旦可以用科学实践去验证的事情,我们只能选择相信科学实践的结果,因为科学追求的是真。

江晓原:您是说,因为有了科学的真,哲学可以退位?我的看法恰恰相反,霍金晚年把很多的领域都还给哲学家了,他的《大设计》里有哲学思考,他发现科学未必能提供真理,他特别强调了我们对于世界图景的描绘,我们并不能知晓外部世界的真相。因为存在着未知,好多科学家最后都会回归到哲学,对未知的事物你可以去想,而思辨的结果并无对错,也无真假。

霍金所主张的“依赖于图像的实在论”,强调了人类有过多种描述外部世界的图像,今天所用的图像也不是终极的。而所有的图像在哲学上都有同等的合理性。

人造出有意识的机器人,不同于人能永生不灭

市教委新闻从业人员宿铭珊:美剧《黑镜》讲述了人肉体死亡之后人的意识会转移到一个永生的空间中。当意识能被模拟后,人类会否将自己的意识进行转移,成为具有意识的永生不灭的机器人呢?

林龙年:在一些国外科幻电影中,常描述了有技术能够实现意识上传。但有一天我们真可以制造出意识机器人,也并不意味着就一定能够传输意识。我们人类具有自我认知能力,这种意识称为自我意识,假如我们能制造出有意识的机器人,并且设置他们以保证自己躯体生存为目标,这时他会以最大的能力去求生,但这与意识的传输还是有很大差别的。

危辉:人类社会发展到今天,有四个悬而未决的基本科学问题:物质的组成、宇宙的起源、生命的进化和意识的奥秘。意识是一种涌现的整体性效应,基于分解与还原策略的科学分析方法可能还不足以解释它产生的原因。因此,还很难讲意识是如何被物理实现的,也就很难讲传输或移植的技术途径。

赫拉利断言“无机人终结有机人”有悖论

工程师柳定毅:《未来简史》作者赫拉利6日在北京的演讲中提到,有机人种将逐步终结,有生之年将看到无机人种搬到火星上。您怎么看?

于海:赫拉利说迄今所有的智能机器都只是智力,而没有人的意识,他说对了;但接下去关于有机人种将被无机人种终结的断言,却是假定无机人有意识,否则如此断言就完全无效了。

是的,人的智力工作看似都能交给人工智能,但智能不代表人的一切意识活动。无机人种或能读懂人的喜好,确切说与智力有关的行为喜好,但它们无心无肺,所以读不懂我们为什么喜为什么怒;读不懂我们为什么喜欢《哈姆雷特》,或《红楼梦》,它更读不懂人的信仰。只要赫拉利相信,硅基“生命”永远无法具有碳基生命的意识,即便硅基算法能搞定大多数碳基人(芸芸众生的习性,多半是程序性的,所以也容易被算法搞定),也无法搞定极少数碳基人,他们的智力和精神不服从任何算法,是他们为算法立法,极而言之,是人类中的发现者创造者发明和创造了AI,他们会不会为硅基生命装上碳基生命的“心”,在技术上是否可能,现在大概无法预知;在伦理上是否可能,只要想一想,这关乎人类自身的存在还是毁灭,我相信未来人类,不会让被自己创造的AI终结自己。

(圆桌对话已刊发于昨日文汇报第8版)