



俞扬, 南京大学计算机科学与技术系副教授, 现为国际人工智能联合大会 IJ-CAI' 15 高级程序委员、IJCAI' 16 宣传主席、IEEE 计算智能协会数据挖掘与大数据分析技术委员会委员、IEEE 计算智能协会演化计算技术委员会委员、中国人工智能学会机器学习专委会委员

人工智能革命“助燃剂”: 机器学习

随着统计建模、基础数学、计算机系统、芯片设计以及应用数学等领域投入的加大, 以及神经科学的不断进步, 机器学习基础科研领域得以快速发展。作为人工智能的分支之一, 机器学习的发展又推动了人工智能的不断进步。机器学习是一种让计算机能够通过经验不断提高自身性能的学科, 可使计算机在未事先明确编程的情况下做出正确反应。在过去 10 年中, 机器学习已经在自动驾驶汽车, 实用语音识别, 有效网络搜索, 以及提高人类基因组认识方面带来大量帮助, 在数据挖掘、自然语言处理、信贷决策、医学诊断、生物信息学、电力监控、网络入侵检测、天气预报、工业控制等领域也已有广泛的应用。

《科技导报》邀请南京大学计算机科学与技术系副教授俞扬解读新一轮人工智能革命下机器学习的发展现状。

人工智能发展浪潮中“技术进步”和“社会进步”不能混淆

“人工智能已经迎来第 3 次发展浪潮, 如何看待这袭来的第 3 次浪潮, 首先不能混淆‘技术进步’和‘社会进步’两者的概念。”俞扬说, 人工智能的概念自 1956 年被提出后, 就快速进入了第 1 阶段的发展时期。到 20 世纪 60 年代,

人工智能程序已经能够自动证明出罗素《数学原理》第 2 章中的大部分定理。20 世纪 70 年代末, 人工智能发展“遇冷”, 一度进入“人工智能之冬”。而后知识的引入使得人工智能进入第 2 次快速发展时期, 在这一阶段, 机器能够发挥其快速推理的作用, “专家系统”的成功应用掀起了人工智能的热浪, 但随着日本五代机计划的破产, “人工智能之冬”重演。直到 20 世纪 90 年代, 机器学习的出现使得从数据中自动提炼知识成为可能, 现如今, 人工智能第 3 次发展浪潮已全面袭来。

俞扬认为, 虽然人工智能迎来了第 3 次发展热潮, 但“人工智能威胁论”尚不成立, 人工智能发展“奇点”也未来临。在谈及人工智能发展时, 他说不能混淆“技术进步”和“社会进步”两者的概念。俞扬举例称, 蒸汽机带来了工业革命, 极大提高了人类生产力, 改变了人类社会。但这巨大进步, 并非是蒸汽机本身的进步, 因为如果依照“奇点”理论推算, 蒸汽机应当迅速发展, 而实际今天的汽车、飞机并没有使用蒸汽机。同时还可以看到, 从蒸汽机到工业革命的过程, 经历了蒸汽机原理的发现、蒸汽机工艺的提升、蒸汽机使用的普及, 以及普及后为人类社会带来的推进。俞扬提到, 如果分开看, 蒸汽机的原理

从发现起就不再变化; 不断被人类打造升级的, 是蒸汽机的工艺, 让蒸汽机效率更高、价格更低、使用更安全方便; 随后蒸汽机的普及使得生产力大幅上升, 从而改变了社会。技术的进步, 可以为人类社会带来深刻的变革, 这是“社会进步”的层面, 而技术本身的发展, 则是“技术进步”的层面, 这两者完全不同。社会关系的跃变, 并不能说明技术出现了“奇点”。

技术的发展是否会出现奇点? 俞扬还是以蒸汽机为例, 他说在弄清了蒸汽机的原理后, 人类不断精化工程工艺, 使最初只能推动锅盖的蒸汽能够驱动轮船。但要注意的是, 原理上的不足, 只有通过发展新的原理来解决。蒸汽机的原理并没有任何变化, 工程工艺的精细化无法克服原理上的先天缺陷, 因此, 尽管蒸汽机有了长足发展, 但在许多地方还是被内燃机取代, 这也就是在原理上有了进步。因此可以观察到技术的进步, 是如图 1 中红色曲线一般“S”型发展, 但如果只看发展的前部分, 就容易中了“奇点”理论的“陷阱”。

俞扬说“奇点”理论中一个有力的“证据”是芯片发展的“摩尔定律”: 英特尔创始人之一戈登·摩尔在 1965 年预言, 半导体芯片上集成的晶体管和电阻数量将每年增加一倍。每年翻番也就

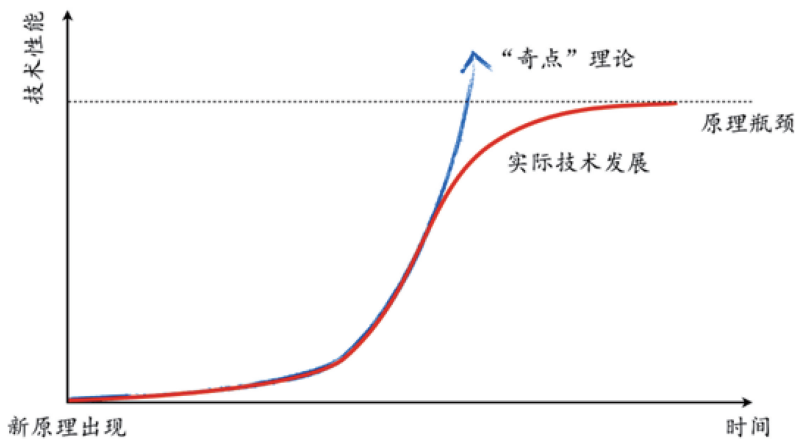


图1 技术进步发展(图片来源:俞扬)

是呈指数增长,如果保持这样的速度,“奇点”很快就会到来。而事实上,1975年摩尔将预言修改为每2年增加一倍,摩尔定律从早期的预言转变为后来的指导原则,每年各芯片厂商都在为实现摩尔定律付出越来越多的努力,而在2016年,新发布的国际半导体技术发展路线图不再以摩尔定律为指导,“摩尔定律失效(the chips are down for Moore's law)”的报道也登上了《Nature》杂志。俞扬还提到,从20世纪90年代到21世纪初,电脑CPU主频迅速增长,如今电脑CPU的主频与几年前并无差异,CPU的发展方向从“高速”变成了“多核”,又在向“低能耗”转变。这些所有的现象,都体现的是“S”型发展,“奇点”从来都没有出现过。

2016年3月谷歌围棋软件AlphaGo战胜李世石时,俞扬说,我们可以期待人工智能的进步能让我们的生活变得更好,同时也需了解到,AlphaGo是站在10年前出现的原理的肩膀上。10年前的原理突破带来今天的技术颠覆,10年后的颠覆也必然建立在今日原理的突破上,而这需要长期基础研究的探索和积累,绝不是一蹴而就。

机器学习“助燃”人工智能革命

“在人工智能60年的发展历程中,机器学习的快速发展起到了非常重要的‘催化剂’作用。”俞扬说,机器学习是机

器(计算机)通过算法,使计算机在存储的大量历史数据中挖掘所需信息,并从中学习规律,进而智能识别新样本或预测未来。机器学习属于人工智能的一个分支,也是目前计算机科学中最活跃的研究分支之一。随着机器学习的快速发展,其应用范围越来越广,越来越多的学者投入到机器学习领域的研究中。2010年图灵奖得主美国Harvard大学L. Valiant教授和2011年图灵奖得主美国UCLA大学J. Pearl教授都是机器学习领域的学者。

在俞扬看来,机器学习可以被划分为“机械学习”、“示教学习”、“类比学习”和“归纳学习”4种类型。计算机计算能力、通讯能力和存储能力的快速发展,也促使人类收集数据能力的显著提高。在对利用数据需求的增加与归纳学习的发展相互促进作用下,自20世

纪80年代以来,归纳学习成为机器学习中被研究最多、应用最广的分支。归纳学习的目标是从个例数据中进行抽象、发现个例背后的规律。以挑西瓜为例,当看到一个根蒂卷曲的西瓜切开是个熟瓜,另一个根蒂青硬的西瓜切开是个生瓜,归纳学习就能从这2个样本中总结出瓜的生熟与根蒂之间的关系,关系存储在机器学习模型中,该模型并能够将归纳出来的关系用来挑选其他西瓜。

这样的归纳学习称为“监督学习”(图2),可以想象为有一个老师给出了一批训练的科目(样本)和每个题目的答案(标记),监督学生(机器)学习,把题目和答案对上。在AlphaGo中也使用到了监督学习:从十余万盘人类对弈的棋谱中学习如何落子,棋盘上的摆子就是样本,这一局棋胜方的落子则是标记。通过监督学习,AlphaGo训练了模仿人类落棋的神经网络。

监督学习需要样本与标记一一对应,而标记通常是由人类专家给出,因此标记的获取非常昂贵,数量相对稀少。归纳学习中除了有监督学习,还包含从没有标记的样本中学习的“无监督学习”,以及介于监督学习和无监督学习之间的“弱监督学习”。强化学习可以看作是一种弱监督学习,它的数据标记需要靠自己探索来获得,往往需要经过多次决策的探索才能获得标记。例如利用强化学习来学习下围棋,可以通过让机器自己跟自己对下来进行,在下棋中产生很多棋局,但这时由于没有监

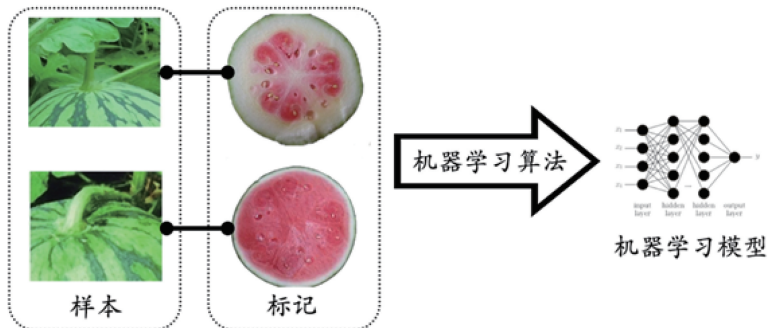


图2 监督学习示意(图片来源:俞扬)

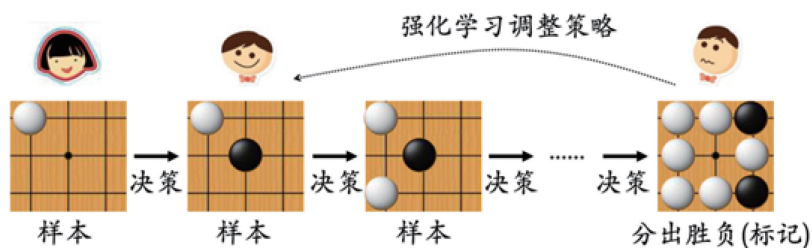


图3 强化学习示意(图片来源:俞扬)

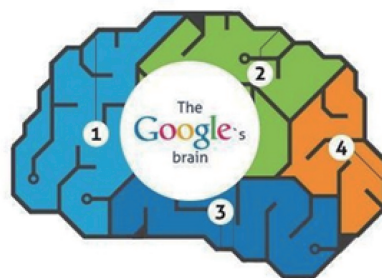


图4 谷歌大脑示意(图片来源:俞扬)

督标记,并不知道每一个棋局应该落哪颗子是正确的,而只有下到分出胜负时,才能倒推之前落子的好坏。这里下棋直到分出胜负就是强化学习中的多次决策,胜负就是强化学习的标记。在AlphaGo使用监督学习训练后,又进一步使用了强化学习来通过自我对弈提高自己的棋力(图3)。

关于机器学习的更详细内容,俞扬推荐了最近出版的周志华教授所著《机器学习》一书。除此之外,俞扬也提到,机器学习不仅在需要“智能”的领域被广泛应用,近年来也逐渐渗透到计算机“底层”研究领域,例如编译器、软件工程和处理器设计方面,同时,计算机其他领域也出现了专门围绕机器学习的发展,例如面向机器学习的处理器芯片、分布式系统、编程语言等等。

目前,许多互联网公司也都看到了机器学习的广阔前景,在机器学习理论及应用中投入了大量人力物力,并初步研发出了实际应用产品。例如,谷歌公司的“Google X”实验室开展的“谷歌大脑”的项目中,学习系统在利用大量数据训练之后,这个系统初步具备了自我

学习能力。作为目前全球最大的学习系统之一,其由1.6万台计算机处理器连接成。自我学习能力的提升,使谷歌图片搜索不再依赖文字描述就可以得到确切的图片,也有助于提供谷歌无人驾驶汽车以及谷歌眼镜的性能(图4)。

关于机器学习模型,深度神经网络模型由于取得了成功应用,最近引起了大量关注。深度神经网络是模拟生物神经元和神经元的连接构建的机器学习模型,然而在俞扬看来,深度神经网络与其说是生物神经网络的模拟,不如说是一种数学模型,目前主流神经网络的学习都依赖具体的数学推导,与生物神经网络的学习相去甚远。神经网络是一种易于操纵的模型,因此可以根据科学家的灵感来构建,并且易于并行处理。然而它也有难以克服的缺点,例如难以把握其学习性能等。虽然最近深度神经网络的关注度增长迅速,历史上同样发生过神经网络火热和冷淡的交替,因此长期来看应当是多种机器学习模型并驾齐驱、优势互补、螺旋上升的发展道路。

机遇与挑战并存

如果我们理解“互联网+”的目的是利用互联网准确获知每一个用户的需求,从而提供高度个性化的产品和服务,那么人工智能技术可以说是不可或缺的。俞扬强调说,如何理解每一个用户的倾向、预测用户需求的发生、匹配最能满足需求的服务等问题,即是发挥人工智能技术的好机会,又是验证技术、提出进一步发展要求的契机。

“人工智能技术近年来有了显著的进步,必然也会迅速普及,应用到生活的方方面面,这也许就是我们正在迈入的‘人工智能+’时代。”俞扬说。国际人工智能联合大会主席、香港科技大学计算机系主任杨强教授也曾提到,“现在中国人工智能的水平和国际几乎没有差距”,在未来的“人工智能+”时代中,中国的人工智能技术的发展和應用都将保持世界前列,并有望引领国际人工智能的发展。

文/祝叶华(《科技导报》编辑部)
(责任编辑 陈广仁)