

人工智能进入后深度学习时代

张钊

(清华大学人工智能研究院, 北京 100084)

摘要: 简要探讨了人工智能的学科概念、进展, 及其未来走向。

关键词: 人工智能; 深度学习; 下一代人工智能

中图分类号: TP181

文献标识码: A

doi: 10.11959/j.issn.2096-6652.201913

Artificial intelligence is entering the post deep-learning era

ZHANG Bo

Institute of Artificial Intelligence, Tsinghua University, Beijing 100084, China

Abstract: Points of views on the concept of artificial intelligence (AI), the progress AI has made, and the trend of AI's future development were presented in this paper.

Key words: artificial intelligence, deep learning, next generation of AI

1 引言

人工智能 (artificial intelligence, AI) 是一门怎样的学科? 目前进展到什么程度? 它又该向何处去? 借《智能科学与技术学报》创刊号出版的机会, 谈一点我的粗浅看法, 供读者参考, 并以此祝贺《智能科学与技术学报》创刊!

AI 尽管有很多定义, 不同的学者对它也有不同的理解, 但是目前较普遍的一种认识是, AI 是计算机科学的一个分支, 它的任务是研究与设计智能体 (intelligent agents)。智能体是指能感知周围环境 (perception), 经理性思考 (rational thinking) 后, 采取行动 (action), 使其达到目标的成功率最大化。由于实现 AI 的主要工具是计算机, 因此很自然地, AI 一直在计算机科学与技术的领域中发展至今。

从 1956 年诞生以来, AI 发展的道路跌宕起伏, 原因其实很简单, 就是我们对于人类的智能 (自然智能) 了解得太少。也就是说, 我们是在对人类大脑工作机制了解很少的情况下, 去研究与实现机器智能的, 这当然是一项很难完成的任务。那么 AI

的学者们如何去完成这项任务? 和其他科学领域一样, 采取的方法是: 大胆假设、小心求证。通过不懈的探索, AI 的研究至今已经经历了两代, 现正进入第三代, 即后深度学习时代。

第一代 AI 是由 AI 奠基者们共同创立的, 最初的思路与假设, 是由认知心理学家提出的, 他们在已有脑科学和心理学知识的基础上提出一种假设, 即物理符号系统假设 (physical symbol system hypothesis, PSSH), 内容如下: A physical symbol system (such as a digital computer) has the necessary and sufficient means for intelligent action^[1]。

基于这个假设, 经过不同领域学者, 特别是计算机科学家的共同努力, 至今有一套建造 AI 系统的理论与方法, 即以知识和经验为基础的推理理论与方法, 通常称之为“符号主义” (symbolism) 学派。回顾 60 余年不平凡的实践, 我们对符号主义的理论与方法又有什么样的结论? 由于人类的理性智能 (行为) 是以知识、经验与推理能力为基础的, 因此符号推理模型适合模拟人类的理性行为。但其中也存在不少困难, 比如, 知识与经验大多数

收稿日期: 2019-03-12; 修回日期: 2019-03-24

通信作者: 张钊, dcszb@tsinghua.edu.cn

不易准确描述，专家的知识与经验十分匮乏且难以获取；人类的知识与推理机制充满不确定性以及存在大量的常识等，这些因素阻碍了符号推理理论与方法的发展与大规模应用，引起了 AI 一段时间里的停滞与挫折，即所谓“AI 的冬天”。不过，进入 21 世纪之后，由于网络与大数据的出现，符号主义的理论和方法得到了一定程度的复苏，其中最具有代表性的工作是 IBM 公司研发的 Watson（沃森）系统，该系统于 2011 年 2 月 14 日在美国电视知识问答竞赛节目危险边缘（Jeopardy）中战胜了前全美冠军 Brad Rutter 和 Ken Jennings。这个系统的成功之处在于使用了大量的数据（原始文本），如百科全书、词典、文学作品和有线电视新闻等，共 4 TB（折合成纸质出版物大约是 2 亿页）。这个系统不仅拥有大量的知识（包括原始文本），还有多种推理机制，能够实现分类推理、空间推理等简单的推理功能。IBM 的工作表明，在人机对话中，如果提的问题是明确的（没有模糊性和二义性），答案是唯一的，领域是有限的——在符合这些条件的情况下，机器可以达到甚至超过人类的水平。这意味着，机器将会在许多商业咨询服务中得到广泛的应用，如人工咨询问答系统等。符号主义在大规模应用上十分有限，除其自身一些缺点（如难以构建、就事论事、难以推广等）之外，还有部分原因是当时计算机的计算能力有限。尽管还存在许多弱点，符号主义依然是实现 AI 的一个重要模型，对于推动 AI 的发展有着重大的意义。因此，在为创建和实现这个学派做出贡献的学者中，先后有 6 名获得了图灵奖，他们分别是：Marvin Minsky（1969 年），John McCarthy（1971 年），Allen Newell、Herbert A. Simon（1975 年），Edward Feigenbaum、Raj Reddy（1994 年），均来自美国。

第二代 AI 是以深度学习为代表的统计学习理论与方法，这种模型的思路来自连接主义（connectionism），而后者是由神经科学家和数学家提出的^[2-3]，它建立在以下的假设基础之上：The central connectionist principle is that mental phenomena can be described by interconnected networks of simple and often uniform units.

经过几十年的徘徊与曲折前进，在各个领域的研究人员，尤其是计算机科学家的努力下，最终形成了一套人工神经网络的理论与算法，使 AI 在 21 世纪初进入了欣欣向荣的深度学习时代。回顾 60

多年的 AI 历史，我们又该如何评价这个时代？人工神经网络模型适用于模拟人类的感知，由于它能够处理和利用大数据，因此获得了广泛的实际应用。比如在给定的数据库（图像库或语音库）下，图像识别和语音识别的准确率可以达到甚至超过人类的水平。尽管深度学习有许多优势，但是通过它构建的 AI 系统往往表现出以下致命的弱点：不可解释性、脆弱性（顽健性很差）、很容易被欺骗和攻击、需要大量的数据等。这些弱点使其只能在有限的场景下使用，如完全信息、确定性信息、静态（或按照确定性规律演化）环境、限定领域等。在发展第二代 AI 中做出重大贡献的 5 名学者先后获得了图灵奖，他们分别是：Leslie G. Valiant（2010 年），Judea Pearl（2011 年），Yoshua Bengio、Geoffrey Hinton、Yann LeCun（2018 年），其中 4 位来自美国，1 位来自加拿大。

深度学习的上述弱点是致命的，是学习机制本身带来的，因此是本质性的。近几年，AI 学者开始思考下一代（即第三代）的 AI。在 2016 中国计算机大会（CNCC2016）上，我们最早提出了“后深度学习时代”的概念，本文中的许多观点均是在那次会议上提出的。2017 年 DARPA（美国国防高级研究计划局）启动可解释人工智能（explainable AI）计划，发展可解释、可信的人工智能技术。2018 年 9 月，DARPA 启动 AI Next（下一代人工智能）战略，提出发展第三波（third wave）人工智能基础理论和技术；同期还启动了 LwLL（learning with less labels）项目，以提升机器学习在数据利用方面的高效性。那么，第三代的 AI 应该是怎样的？大体说来，有以下几点：建立可解释、顽健的人工智能新理论与新方法；发展安全与可信的人工智能新技术；推动人工智能的创新应用。通过什么途径才能达到这个目标？大体说来，有两个途径，一个是与脑科学，特别是神经科学结合，探索新的模型和原理；另一个是，把知识驱动与数据驱动结合起来，发挥各自的优点，克服各自的弱点，达到互补的目的，以实现向后深度学习时代的转变。

2 结束语

当 AI 进入新的时代——后深度学习时代，各国都将处于同一起跑线上，谁能跑到最前面并引领 AI 的发展，完全取决于其创新能力，特别是原始创新能力。因此加强基础理论研究至关重要，因为只

有理论上取得进展，才能带来关键技术的突破，从而引发新一轮的创新应用，充分把握新时代给予我们的机遇与挑战。

参考文献：

- [1] NEWELL A, SIMON H A. Computer science as empirical inquiry: symbols and search[J]. Communications of the ACM, 1976, 19(3): 113-126.
- [2] WARREN M, PITTS W. A logical calculus of ideas immanent in nervous activity[J]. Bulletin of Mathematical Biophysics, 1943, 5 (4): 115-133.

- [3] DONALD H. The organization of behavior[M]. New York: Wiley. 1949.

[作者简介]



张钹（1935— ），男，福建福清人，中国科学院院士，清华大学教授、博士生导师，清华大学人工智能研究院院长，主要研究方向为人工智能、人工神经网络等。