

· 科学论坛 ·

类脑智能与深度学习的几个问题与思考*

焦李成¹ 杨淑媛¹ 韩军伟^{2**}

(1. 西安电子科技大学 人工智能学院, 西安 710071;
2. 西北工业大学 信息融合技术教育部重点实验室, 西安 710072)

[摘要] 得益于数据规模与机器算力的增长,深度学习技术正取得空前繁荣。然而,深度学习擅长在预设的封闭环境,为特定任务找到有用的数据表示,在结构、处理和功能上距离人类智能仍有较大差距。类脑智能旨在模拟人脑神经元的运行机制、感知模式与认知机理,借助机器强大的信息整合、搜索、计算等能力,以软硬件联合的智能新形态构造接近人类水平的智能机器,是未来人工智能的发展方向。基于第216期“双清论坛”,本文将分析人工智能与深度学习发展现状与局限,并从认知建模、模块装配、意识先验、自主演化、协同学习几方面分析未来类脑智能的可能发展方向。

[关键词] 类脑智能;类脑认知;模块化;意识先验;自主演化;协同学习

人工智能(Artificial Intelligence, AI)起源于控制论中的反馈机制,自1956年达特茅斯会议人工智能的概念被正式提出,至今已有六十余年的历史。经过半个多世纪的演进,人工智能已迈入了新的发展阶段。目前,得益于互联网和各种智能设备产生的海量数据、大幅提升的机器运算能力以及最近几十年机器学习领域的突破,以深度学习为代表的现代人工智能技术在许多应用中的性能都取得了突破性进展^[1-9]。例如,在语音识别中机器的错误率已低至6%^[8],在人脸识别中机器的错误率甚至低于0.2%,达到了远高于人类的水平。2016年,AlphaGo战胜李世石、无人驾驶上路,2017年,AlphaGo零封柯洁,更是吸引了学术界与工业界的空前热情。2018年,DeepMind AlphaGo团队被人工智能领域顶会IJCAI授予首个Marvin Minsky奖。2019年1月,AlphaStar在《星际争霸2》人机大战中大胜人类玩家,再一次证明了人工智能的强大能力。目前以深度学习为代表的人工智能技术在计算机视觉、语音识别、自然语言理解、机器博弈等领域都形成了不小的冲击^[1-11],其产业化也成了世界主要发达国家提升国家竞争力、维护国家安全的重



焦李成 教授,博士生导师。现任西安电子科技大学计算机科学与技术学部主任、智能感知与图像理解教育部重点实验室主任、教育部科技委学部委员、教育部人工智能科技创新专家组专家、IET西安分会主席、IEEE西安分会奖励委员会主席、IEEE TGRS 副主编、IEEE/IET/CAA/CAA/CIE Fellow。主要研究方向为智能感知与计算、图像理解与目标识别、深度学习与类脑计算,研究成果获青年科技奖、国家自然科学基金二等奖及省部级一等奖以上科技奖励十余项。



韩军伟 2003年获得西北工业大学博士学位,现任西北工业大学教授。长期从事人工智能、模式识别等领域的研究。曾获陕西省科学技术一等奖等。

大战略。在我国《中国制造2025》《机器人产业发展规划(2016—2020年)》以及《“互联网+”人工智能三年行动实施方案》^[12]中,人工智能都被列入核心发展对象。

收稿日期:2019-06-04;修回日期:2019-10-18

* 本文根据第198期“双清论坛”讨论内容整理。

** 通信作者,Email:hanjw@nsfc.gov.cn

1 深度学习

尽管深度学习已经在一些特定任务如人脸识别、语音识别、场景分类等应用中取得了很大飞跃^[13-17],达到逼近乃至超越人类的水平,相关技术也已逐渐成熟落地到生活的多个方面。作为现阶段人工智能的主流技术,深度学习模型的工作机理为:依赖人类筛选和准备的训练样例,基于多层非线性变换来准确刻画和记忆数据。因此,现有深度模型擅长工作于封闭环境,为特定任务找到有用的数据表示^[18]。然而,已有的深度学习模型离多领域通用的“强人工智能”的实现还有不少差距,在高级感知、推理、交互,以及渐进学习、自主学习和交互式学习等方面还存在明显欠缺:

(1) 模型阐释难题。现有深度神经网络模型能将图像、文字、语音等数据映射至数学上的高维矢量空间,使不同事物的矢量距离远近反映出其重要区别,但获得的每个特征与总体决策之间的关系如何,以及什么是提升深度学习系统的核心等关键问题,仍不具备可解释性^[19]。这直接导致了深度学习在军事、医疗等一些解释性优先的应用中受限。

(2) 模型认知缺陷。目前深度神经网络模仿人脑行为的方式仍较简单,存在模型结构单一、仅能应用于某特定领域、缺乏环境自适应性等问题。尽管在感知智能方面已有众多成果^[20],但在知识抽象、自主学习、渐进学习、直觉与遗忘等认知智能层面的能力较弱。如何在对新领域和新环境一无所知的情况下,学习并巩固所发现的内容,建立可复用的复杂模块化组件库,构建关于新领域的知识体系,将感知功能提升到认知与意识水平,是当前深度学习面临的难题与瓶颈。

(3) 模型学习瓶颈。深度模型大多基于优化高度非凸性的经验风险损失和扩展的随机梯度优化算法来学习网络参数。学习算法的一致性收敛、稳定性、经验风险的泛化、梯度和驻点的属性,以及网络深度、宽度、大小等架构参数和网络权重参数等对学习算法收敛的影响等一系列问题,至今仍不明确。由于其深层的结构与庞大的参数,导致模型的学习过程非常微妙,并且需要精细、耗时、耗力的超参数搜索、调整和凑试,这些学习中的瓶颈问题使其难以进行高风险与高可靠地应用。

(4) 模型资源限制。现有深度神经网络模型的成功大多依赖大量资源消耗和大规模的标记数

据,导致深度学习的训练与实际应用需要耗费大量的存储资源与计算资源。谷歌公布的 AutoML 模型需要超过 800 个 GPU 全天候运行数周,来获得较好的网络架构与参数。此外,现有深度模型的全局更新机制导致其配置与计算成本非常昂贵,无法在开放和泛在(如移动计算)环境下获得实时的处理结果,这些都造成了模型学习过程中的资源限制。

2 未来类脑智能发展思考

作为“面向特定应用”的弱人工智能技术,深度学习距离多领域通用且具备多重智能行为的“强人工智能”目标仍有不少差距。当前人工智能正处在从“弱人工智能”到“强人工智能”的质变过程。类脑智能旨在模拟人类大脑神经元运行机制、感知模式与认知机理,借助机器强大的信息整合、搜索、计算等能力,软硬件联合以形成接近人类智能水平的智能机器,是未来人工智能的发展方向。一方面,类脑智能强调计算机需要具备像人一样的多重智能行为,包括感知、决策、推理与规划,以及交流、沟通等。另一方面,类脑智能强调以软硬件协同实现智能的新形态,期望基于神经形态工程,打破“冯诺依曼”架构束缚、研究适于实时处理非结构化信息、具有学习能力的超低功耗新型计算方式。

类脑智能不但要从结构上模仿大脑,而且还要从神经元和突触的模型上模仿大脑,其目标是以物理的形态实现大脑功能与其结构的对应,并以并行分布和自主学习的方式构成大规模神经网络计算系统。没有一流的计算神经科学,就不可能获得一流的类脑智能技术创新。计算神经学科侧重对脑系统的理论建模和计算仿真,通过该学科严谨的理论、建模和统计分析,从而对复杂的、非线性的大脑系统进行深入剖析和建模分析,尤其是大尺度的认知功能脑系统计算模拟、跨层次机制分析,计算神经科学搭建起了计算机科学和脑科学之间的桥梁,这将为类脑智能、脑机接口和人工智能的发展起到关键作用。通过计算神经科学与计算机科学的密切合作,完成类脑智能及脑科学的重大前沿问题的联合攻关。目前世界各主要发达国家都在积极推动领域通用的类脑智能的发展。本文将从以下几个方面进行讨论:

(1) 认知建模。脑机理认知建模是类脑智能实现的基础。目前人类对于大脑认识还极为有限,神经元与其连接的多样性、脑网络结构的复杂性、多脑

区间的协同性给认知建模带来困难。过去十年间国际上兴起了脑科学研究的热潮,发达国家和我国也纷纷推出大型脑研究计划^[21-24]。脑与神经科学、认知科学的进展使得在脑区、神经簇、神经微环路、神经元等不同尺度观测的各种认知任务中,获取脑组织的部分活动相关数据已成为可能。因此,未来借助认知研究成果,构建认知脑计算模型的多尺度(神经元、突触、神经微环路、皮质柱、脑区)计算组件和多脑区协同模型,其中包括类脑的多尺度前馈、反馈、模块化、协同计算模型等;模拟认知/智能行为的类脑学习机制(如多模态协同与联想的自主学习机制,概念形成、交互式学习、环境自适应的机制等),将对构建类脑智能计算模型提供发展契机。

(2) 模块装配。脑的主要结构特征之一就是模块化^[25]。模块化设计早已被计算机科学所采用^[26]。在设计中,问题首先解决分成几个固定的部分(子问题),每个计算模块(子程序)只负责处理其中的一个。模块化设计的优势显而易见,但面对一系列具体问题,应该如何最有效率地划分子问题,这本身是一个困难的任务。因而,这可能是我们能够向大脑学习的重要知识之一。

真实的大脑是模块化设计的一个范例,每个脑区或亚区负责一个信息处理的环节,而具体的模块划分是经过漫长自然选择加以优化的结果,已经适应了高效处理真实世界的实际问题。因此,在发展类脑智能技术的过程中,可以向脑学习如何更好地实现算法设计的模块化。一方面,使算法设计大大简化,易于调试、易于修改,可以逐步完善并增加功能;另一方面,因为表面上看起来不一样的问题往往能分解成相似的子问题,这使得模块可以重复利用,大大提高了效率并使得高度简并的系统能够胜任复杂多样的任务。此外,如果以可了解(学习能力可知、计算资源可知、典型任务性能可知等)和可重用的模块化组件生成新模型来代替深度网络黑箱,分析其理论基础与群体智能效应,有望解决深度学习阐释难题。

(3) 意识先验。意识是人类智能与现有机器智能的重要区别。Bengio 将意识看作人脑在某一特定时刻的想法,它来自人类观察,包含了足够多的知识,但仅能容纳少数几个要素^[27]。因此可以用一个低维向量表示该状态,称为有意识状态(Conscious State)。相反的,人脑中的所有内容则可看作是一个高维但稀疏的向量,称为无意识状态

(Unconscious State)。意识先验能够从无意识状态中提取出意识状态的少量要素,因此在认知建模过程中可以使用意识先验,将有望将符号主义与连接主义结合,在面临复杂任务时做出更接近人类的预测结果。

(4) 自主演化。人类具有不断自我学习的能力,能迅速改变自己以适应不断变化的外部环境。该能力来源于人脑中除了基本的兴奋与抑制性的神经递质外,还有众多的神经调质^[27],其作用是根据当前的环境与行为目标随时动态调节大范围神经网络的行为,使其能够胜任复杂多变的情况,实现千差万别的任务。近年来对于介观及宏观脑网络动态活动规律的研究发现,脑网络可能自发地组织于一个“临界”状态附近,这一状态使得信息的存储、传递和处理都能实现最优化^[28]。重要的是,通过对这一状态的微调,可以迅速调节网络功能,从而适应不同任务的要求。

面向开放的真实自然场景,类脑智能模型应具备对已知模式的学习能力,未知模式的探索能力,以及对知识、结构和参数的自主演化与进化学习能力,以自动处理“没见过”的任务。具体来说:1) 模型应能够进行任务的自我演化;2) 模型应能够进行计算单元的自主演化;3) 模型应能够进行结构的自主演化;4) 模型应能够进行学习机制的自我演化。对于神经调质以及网络状态调控等原理的借鉴,有望对设计具备自主演化能力的类脑智能模型提供启示。

(5) 协同学习。知觉及思维的产生正是大脑复杂的神经网络庞大系统中各子系统的协同综合作用的结果。从微观层面上,生物神经元的结构和突触的类型、数目等在不同脑区中具有较大差异,且能够根据任务的复杂性实现结构和功能的动态适应^[30]。从介观层面上,连接模式与随机性网络背景噪声的有效融合,使得生物神经网络在保持了特定网络功能的同时,兼顾了动态网络可塑性。从宏观层面上,不同脑区之间的协同使得高度智能的类人认知功能得以实现。融合来自不同脑区的信号,从而使对客观对象的认识更为全面^[29]。

大脑的协同机制为构建更高效、更实用的类脑模型提供了生物基础,即以大脑工作机理为基础,深入分析微观、介观、宏观多尺度脑结构,实现协同学习的类脑智能计算模型。应重点突破的创新点包括:1) 跨尺度机制的融合:需要重点关注人脑如何协同不同尺度的计算组件,进行动态认知环路的组

织,完成不同的认知任务,并从此受到启发,构建多尺度的类脑模型。2) 多功能模块、多功能区域的协同学习:借鉴脑神经系统与自主神经系统(受中枢影响,不受中枢控制,是脊椎动物的末梢神经系统),脑不同功能区域(如大脑与小脑),脑不同任务区域(如视觉皮层和听觉皮层)的协同机制,构建高度协同的类脑模型。

类脑智能势必会推动人工智能技术向着更自动化的程度发展,这将给未来人类的决策产生巨大的影响,由此带来的人工智能伦理和监管问题日益凸显。鉴于此,在类脑智能技术研究中,从模型、数据和人本角度出发,构建“值得信赖的类脑智能”。1) 类脑模型的可解释性:提高人工智能的透明度是当前 AI 领域的共识,尤其是针对军事、医疗等应用领域。构建可解释性的类脑模型,使得类脑系统的功能目的、算法模型及处理过程更加透明,从而减少潜在的不安全问题,确保类脑智能的安全、稳健、合法和符合伦理。2) 决策数据集多样性和平衡性:人工智能决策极大地依赖于大数据的驱动,因此在类脑模型的训练中,尽可能公平地选取足够品类和数量的标签数据,保持决策数据集的多样性和平衡性,以减少数据集的不平衡性和不完备性,从而构建出更为客观全面的类脑模型。3) 人本原则,加强监管及问责:人工智能始终围绕“人”展开,由人类创造,且终究成为人类的工具,对人工智能的治理也要回归人类本身。全社会需要协同并肩,共同制定人工智能的伦理规范和应用准则,加强决策数据集的公平、可靠、隐私和保障,减少人为偏见的存在。当人工智能系统出现不好的结果,不能一味的将责任推脱给人工智能系统,而应该问责法人主体。

3 结 语

本文从类脑智能研究的现状出发,分析了当前类脑智能的优势和不足,进而从未来人工智能需求的角度,阐述了对类脑智能研究的几点思考。类脑智能研究不能脱离其他学科单独存在,应和计算神经科学、生物脑科学、计算机科学、材料学和数学等领域紧密联系,以生物脑工作机制为启发,多学科融合实现基于类脑机制的下一代智能系统。

参 考 文 献

[1] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.

- [2] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In CVPR, 2015.
- [3] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In CVPR, 2015.
- [4] Huang G, Liu Z, Laurens V D M, et al. Densely connected convolutional networks. In CVPR, 2017.
- [5] Girshick R. Fast R-CNN. In ICCV, 2015.
- [6] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. International Conference on Neural Information Processing Systems, 2015.
- [7] Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search. Nature, 2016, 529(7587): 484—489.
- [8] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine, 2012, 29(6): 82—97.
- [9] 焦李成, 杨淑媛, 刘芳, 等. 神经网络七十年: 回顾与展望. 计算机学报, 2016, 39(8): 1697—1716.
- [10] Howard J, Ruder S. Universal language model fine-tuning for text classification. In NAACL, 2018, v1: 328—339.
- [11] 焦李成, 赵进, 杨淑媛, 等. 深度学习、优化与识别. 北京: 清华大学出版社, 2017, 240—360.
- [12] 徐贵宝. “互联网+”人工智能技术发展策略解析. 世界电信, 2016(3): 71—75.
- [13] Zhang KP, Zhang ZP, Li ZF, et al. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 2016, 23(10): 1499—1503.
- [14] Li Y, Ai HZ, Yamashita T, et al. Tracking in low frame rate video: a cascade particle filter with discriminative observers of different life spans. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(10): 1728—1740.
- [15] Wen Y, Zhang K, Li Z, et al. A discriminative feature learning approach for deep face recognition. In ECCV, 2016.
- [16] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding. In CVPR, 2016.
- [17] Kim S, Seltzer ML. Towards language-universal End-to-End speech recognition. In IEEE ICASSP, 2018.
- [18] 焦李成, 赵进, 杨淑媛, 等. 稀疏认知学习、计算与识别的研究进展. 计算机学报, 2016, 39(4): 195—212.
- [19] Marcus G. Deep learning: A critical appraisal. In arXiv: 1801.00631, 2018.
- [20] 焦李成, 尚荣华, 刘芳, 等. 稀疏学习、分类与识别. 北京: 科学出版社, 2017.

- [21] Bargmann CI, Newsome WT. The brain research through advancing innovative neurotechnologies (BRAIN) initiative and neurology. *JAMA Neurology*, 2014, 71(6): 675.
- [22] Amunts K, Ebell C, Muller J, et al. The Human Brain Project: Creating a European Research Infrastructure to Decode the Human Brain. *Neuron*, 2016, 92(3): 574—581.
- [23] 杨雄里. 为中国脑计划呐喊. *中国科学: 生命科学*, 2016 (2).
- [24] Okano H, Sasaki E, Yamamori T, et al. Brain/MINDS: a Japanese National Brain Project for Marmoset Neuroscience. *Neuron*, 2016, 92(3): 582—590.
- [25] Bertolero MA, Yeo BTT, Bassett DS, et al. A mechanistic model of connector hubs, modularity and cognition. *Nature Human Behaviour*, 2018, 2(10): 765—777.
- [26] 余山. 从脑网络到人工智能——类脑计算的机遇与挑战. *科技导报*, 2016, 34(7): 75—77.
- [27] Bengio Y. The Consciousness Prior. In arXiv: 1709.08568, 2017.
- [28] Shew WL, Plenz D. The functional benefits of criticality in the Cortex. *Neuroscientist*, 2013, 19(1): 88—100.
- [29] 徐波, 刘成林, 曾毅. 类脑智能研究现状与发展思考. *中国科学院院刊*, 2016, 31(7): 793—802.

Thoughts and prospects of brain-inspired intelligence

Jiao Licheng¹ Yang Shuyuan¹ Han Junwei²

(1. *School of Artificial Intelligence, Xidian University, Xi'an 710071;*

2. *Northwestern Polytechnical University, the Key Laboratory of Information Fusion Technology, Ministry of Education, Xi'an 710072)*

Abstract Thanks to the growth of data scale and machine power, deep learning technology has achieved unprecedented prosperity. However, most of the available deep neural networks are good at providing useful data representation for a specific task in closed environment, which has remarkable difference with real human brain. Brain-like intelligence aims to simulate the processing, perception and cognitive mechanism of human brain neurons, and establish intelligent machines close to human level in a new form including both software and hardware, via the utilization of computers with powerful calculation capabilities. Now brain-like intelligence is guiding the direction of Artificial Intelligence (AI) research works. In this paper, we reviewed the current situation of AI and analyzed the limitations of deep neural networks. Meanwhile, some possible development directions of brain-like intelligence, including cognitive modeling, module assembly, conscious priority, autonomous evolution and collaborative learning, were also discussed.

Key words brain-like intelligence; brain-like cognition; modularization; awareness priori; autonomous evolution; collaborative learning