

认知鸿沟？人工智能该何去何从

李 玲, 熊韦锐*

重庆师范大学教育科学学院, 重庆

收稿日期: 2024年10月23日; 录用日期: 2024年12月9日; 发布日期: 2024年12月23日

摘 要

人工智能是新一轮科技革命和产业变革的重要驱动力量。然而, 人工智能在解决问题中的失误带来的不利后果常常是人类无力招架的。如何优化人工智能的问题解决能力对于其发展至关重要。因此在前人积累和理论分析基础上, 拟从脑科学、心理学研究成果探析生物智能与人工智能的认知差别。认知方式的差异意味着个体对于解决问题的方式不同, 因而人类认知方式是人工智能突破认知鸿沟实现高效正确解决问题的有效途径。本文总结了脑科学、心理学研究成果促使人工智能算法更新的历史, 归纳了顿悟对人工智能算法进步的指导意义, 并提出将顿悟体验纳入人工智能算法模型更迭参考。本文对于探析人工智能认知进步及其算法更新具有较好现实意义。

关键词

人工智能, 脑科学, 顿悟, 顿悟体验

A Cognitive Gap? Where Should AI Go from Here

Ling Li, Weirui Xiong*

School of Educational Science, Chongqing Normal University, Chongqing

Received: Oct. 23rd, 2024; accepted: Dec. 9th, 2024; published: Dec. 23rd, 2024

Abstract

Artificial intelligence is an important driving force for a new round of scientific and technological revolution and industrial transformation. However, AI's mistakes in solving problems often have adverse consequences that humans are unable to cope with. How to optimize the problem-solving ability of AI is crucial to its development. Therefore, on the basis of previous accumulation and theoretical analysis, this paper intends to explore the cognitive differences between biological

*通讯作者。

intelligence and artificial intelligence from the research results of brain science and psychology. The difference of cognitive style means that individuals have different ways to solve problems. Therefore, human cognitive style is an effective way for artificial intelligence to break through the cognitive gap and realize efficient and correct problem solving. This paper summarizes the history of brain science and psychology research results to promote the updating of artificial intelligence algorithms, summarizes the guiding significance of insight to the progress of artificial intelligence algorithms, and proposes to incorporate insight experience into the reference of artificial intelligence algorithm model. This paper is of great practical significance for exploring the cognitive progress of artificial intelligence and its algorithm updating.

Keywords

Artificial Intelligence, Brain Science, Insight, Insight Experience

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

当前世界百年难遇之大变局,研究人工智能对于促进经济发展、提升生活质量、优化公共服务,应对全球挑战都具有重大意义。人工智能(AI)的最新进展已经达到了在分类任务中超越人类的速度和准确性。这些能力使 AI 成为许多以分类任务为核心的人类活动的可行替代品,包括制造业、健康、金融和服务业;并且 AI 在图像、语音识别以及行为预测等定义良好的领域以及基于这些领域的服务方面表现出色(朱毅鑫等, 2020)。此类基础服务工作中的基本机械和分析任务不需要有意识的识别模式;与黑猩猩、猫科动物和人类等自然诞生的生物智能系统相比,所有这些 AI 能力都相当原始。因为大多数 AI 本质上是超高效的分类或回归算法,针对特定任务进行优化训练;他们学习分类标签以及回归结果,并使用该训练来实现预测或分类的指定目标(Aitken et al., 2020)。然而时代发展的必然趋势, AI 需要进阶完成有同理心的、更复杂的、更多以人为本的服务任务,发展类似于人类的创造力和同理心等能力成为先决条件。顿悟是对复杂的知觉情形或问题突然有了清晰的理解和解决,且往往伴随啊哈体验(沈汪兵等, 2013)。如何使 AI 具备这一能力,并从算法层面付诸实践,使 AI 具有类似于顿悟的更快速、创造性的问题解决能力,整理人类顿悟产生的认知神经机制成为必要,对 AI 算法更迭具有重要参考意义。

2. 受脑科学影响的 AI

自 AI 诞生以来, AI 与脑科学之间就存在联系,无论是 AI 的模型换代还是算法更迭,脑科学的研究成果都起到了重大作用。AI 是能代替生物智能即人类、动物执行任务且具有感知、识别、决策和控制等功能的计算机系统(Cooper, 2005)。脑科学是研究人、动物和机器的认知与智能的本质和规律的科学(张旭, 2024)。因此,将脑科学的研究成果运用于 AI 算法创新可以取得更高效的进展。

人工神经网络模型(Artificial Neural Network, ANNs)是模仿人类神经网络行为特征进行信息处理的算法数学模型(Abdolrasol et al., 2021)。这是史上最早、最成功的 AI 模型之一,而它的产生正是受到了 20 世纪初神经系统中神经元之间存在联系这一脑科学发现的启发。使 AI 不断适应外界,可塑性得到增强的 Hebbian 学习算法(Sabharwal & Selman, 2011)也是受到了受生物神经系统力学的启发。事实证明解析大脑与神经系统,会启迪 AI 理论与技术。例如优于现有所有图像检索算法的卷积神经网络的产生(Miller, 2003)是受到猫大脑视觉处理系统进行卷积操作(Hubel & Wiesel, 1962)这一研究发现启迪。深度学习解决了很

多复杂的模式识别难题(陈先昌, 2014), 是使机器模仿视听和思考等人类的活动达成。此外, 人工神经网络和深度学习的关键组成部分——反向传播算法, 也受到了神经系统的微观结构和生物大脑的神经系统是通过一个学习过程逐步被调整的, 其目的是最小化误差和最大化输出回报(Scolari, Seidl-Rathkopf & Kastner, 2015)这一成果的启发。

综上, 在 AI 算法中, 生物智能的脑科学研究成果贡献了不可或缺的力量, 一定意义上讲, AI 算法是对生物智能大脑运转机制的模仿。

3. 更“智能”的 AI——心理学的贡献

相较于脑科学研究, 心理学以不同的视角研究生物生理活动的心理机制与生理基础, 为 AI 进步贡献了属于心理学的力量。正如信息加工理论(Information Processing Theory)将人类认知过程与计算机的信息处理视为类同, 这种理论强调了输入、存储和输出在认知发展中的重要性, 而 AI 正是通过算法实现类似的认知过程。一部分研究者认为可以用人类思维过程作为改进计算机处理数据的程序的指南(傅小兰, 1989)。还有一些研究者, 最典型代表的是 Simon 和 Newell 用计算机加工模拟人类问题解决过程(Gugerty, 2006)。在这种理论视角下, 许多心理学研究成果被运用到 AI 发展中。

20 世纪 70 年代工作记忆这一心理学概念被提出(Baddeley, 1981), AI 研究人员试图将记忆模块纳入机器学习模型, 如长短期记忆(Long Short-Term Memory, LSTM)神经网络, 作为一种有记忆性的预测型算法, 具有学习长期依赖关系的能力(Muzaffar & Afshari, 2019)。而工作记忆模块使机器学习模型可以执行复杂的推理和工作任务(Santoro et al., 2016)。AI 需要学习大量样本数据集才能准确执行已经训练过的特定任务。直到双光子显微技术出现, 揭示了神经系统在学习新任务时是如何通过控制神经元的生长来记住之前的任务的(Kirkpatrick et al., 2017)。受到该研究的启发, EWC (Elastic Weight Consolidation)学习算法被提出并被用于深度神经网络的研究, 其使得深度学习中的连续学习成为可能。诸如此类的还有借助人脑信息处理机制, 用脉冲信号表示时序信息的脉冲神经网络(Spiking Neural Network, SNN)计算, 可以有效地处理大量的非结构化数据, 如图像、声音和文本等(张旭, 2024)。由此可见 AI 的进步与发展, 从心理学视角对人类大脑及神经机制的研究是关键。

AI 的进步不仅受益于心理学实证研究的成果, 心理学理论对 AI 算法进步与实际完成任务能力的提升同等具有重要的指导意义。人类视觉具有轮廓感知修复能力, 如何使计算机具有这种能力一直是计算机视觉面临的难题之一。而采用格式塔心理学的完形法提出 T 型节点组合的约束条件, 使平滑轮廓遮挡和角点轮廓遮挡这两种类型的轮廓修复方法对于自然场景的复杂角点轮廓修复也行之有效(张桂梅, 刘丕玉, 2013)。多策略生活选择算法 MSLCBO, 也是基于心理学中的“贝勃规律”、“关系场”与“投射效应”理论的启发, 该算法在高维测试函数和工程应用问题的解决上都得到了提升(刘勇, 胡学敏, 陈卓, 2024)。融入了教育心理学的 SBO-EP 算法(SBO Based On Educational Psychology, SBO-EP), 参考了发展区理论上、成就动机理论, 弥补了原有 SBO (School Based Optimization)算法存在的寻优精度低、全局搜索能力弱等缺陷, SBO-EP 算法性能更强、收敛速度更快且具有更高的稳定性, 在优化方面也具有更强的竞争力(张雨婷, 刘勇, 2022)。

由此可见, AI 的发展极大程度上受到脑科学研究以及心理学研究的影响, 研究生物智能的脑机制以及生理机制, 以及心理学的相关理论与发生发展机制对于启发人工智能算法进步有着不可忽视的作用。

4. 顿悟——AI 的未来

目前 AI 问题解决仍依赖于对 AI 进行大量数据训练来解决一项单一的任务, 且缺乏对物理世界与社交世界运作常识的理解(朱毅鑫等, 2020)。显然这样的 AI 是不够“智能”的。与此相反, 人类早在婴儿

时期, 就能迅速有效地感知因果关系(Scholl & Nakayama, 2004), 可以使用较少的经验应对许多场景。这给当前 AI 问题解决带来了巨大挑战。解决此挑战的一种方法是参考人类问题解决的方法——顿悟。

顿悟作为创造的一种重要形式, 一直是研究的热点。很多人在解决问题时都有过百思不得其解, 突然灵光乍现得到问题答案的体验, 即顿悟体验。有研究者提出顿悟的核心特征正是顿悟体验(insight experience), 即“啊哈”体验(Danek & Salvi, 2020)。这种突然性强调了顿悟的不连续性本质。个体能够在瞬间超越之前的认知框架, 洞察问题的本质或找到解决问题的新途径。也有研究者 Luo & Knoblich (2007) 强调, 顿悟问题解决通常是通过问题元素的重新组合产生新的认知视角, 从而解决问题。这种突然性的、创新的问题解决方式为当前 AI 提升问题解决能力与速度提供了新思路以及可行性前进方向。研究探明顿悟的认知神经生理发生机制对 AI 算法的进步具有实践指导意义。

谭钢等人(2013)以 CRAT 任务为材料借助 ERP 技术探究自发顿悟与非顿悟之间的脑电差异, 结果显示可能与打破心理定势有关; 以及可能反映了形成新联想连接的过程。Jung-Beeman 等人(2004)使用 fMRI 技术采用 RAT 问题探究顿悟与非顿悟在神经机制上的差异, 结果显示以顿悟的方式解决问题能够引起 ACC 和双侧杏仁核的激活。林家宝等人(2018)使用了采用 fMRI 技术来探究顿悟体验产生时刻的神经关联, 结果探明了顿悟问题解决与大脑自发的神经活动的关联性。结合以上研究结果来说, 顿悟体验时刻与轻松和快乐的感觉、个体对决定的确定性编码、粗略的语义编码等均有关联。Ohlsson (2011)更是明确提出创造性的问题解决存在一个顿悟序列(Insight Sequence), 即 1 搜索、2 僵局、3 顿悟、4 结果。如何将这种顿悟序列运用转化到 AI 问题解决中, 需要心理学研究者进一步研究顿悟产生的认知神经机制。例如邢强等(2017)对顿悟研究的进一步研究发现, 顿悟条件诱发的生理反应揭示了思维僵局的早期觉察和旧思路的僵局打破过程。此类研究需要更多。

5. 总结

当下社会飞速发展, 对 AI 性能以及智能化的要求越来越高, 但当前 AI 算法已经不能满足, 新算法的出现已经是必然结果。已有研究者提出各式各样的设想, 但如何开发符合的模型及算法, 仍需努力。过往经验表明, 人类智能的心理机能发生机制与生理基础的研究成果极大程度地推动了 AI 范式革命与模型、算法的更新迭代。站在当今这个进化的节点上, 对人类顿悟生理机制研究或将成为 AI 新进展的突破口, 研究探明顿悟的生理发生机制具有重要的参考意义。

参考文献

- 陈先昌(2014). 基于卷积神经网络的深度学习算法与应用研究. 硕士学位论文, 杭州: 浙江工商大学.
- 傅小兰(1989). 思维的信息加工理论和计算机模拟方法. *心理学动态*, (2), 16-20, 25.
- 林家宝, 崔璇, 戴晓颖, 陈雅珏, 莫雷(2018). 静息状态的大脑信号可预测顿悟. 见 *第二十一届全国心理学学术会议摘要集*(pp. 456-457). 华南师范大学心理学院.
- 刘勇, 胡学敏, 陈卓(2024). 基于心理学理论的多策略生活选择算法. *智能计算机与应用*, 14(5), 10-18.
- 沈汪兵, 刘昌, 袁媛, 张小将, 罗劲(2013). 顿悟类问题解决中思维僵局的动态时间特性. *中国科学: 生命科学*, 43(3), 254-262.
- 谭钢, 侯玉玲, 曹智鹏(2013). 遥远距离联想任务脑机制的 ERP 研究. *心理与行为研究*, 11(6), 746-751.
- 邢强, 孙海龙, 占丹玲, 胡婧, 刘凯(2017). 执行功能对言语顿悟问题解决的影响: 基于行为与 ERPs 的研究. *心理学报*, 49(7), 909-919.
- 张旭(2024). 智能时代的脑科学与类脑智能研究. *中国科学院院刊*, 39(5), 840-850.
- 张雨婷, 刘勇(2022). 融入教育心理学的 SBO 算法. *计算机应用研究*, 39(9), 2631-2639.
- 张桂梅, 刘丕玉(2013). 基于格式塔心理学和 Euler spiral 的轮廓修复算法. *计算机应用研究*, 30(8), 2534-2537, 2560.

- 朱毅鑫, 高涛, 范丽凤, 黄思远, 等(2020). “暗”, 不止于“深”——迈向认知智能与类人常识的范式转换. *Engineering*, 6(3), 230-306.
- Abdolrasol, M. G. M., Hussain, S. M. S., Ustun, T. S., Sarker, M. R., Hannan, M. A., Mohamed, R. et al. (2021). Artificial Neural Networks Based Optimization Techniques: A Review. *Electronics*, 10, Article 2689. <https://doi.org/10.3390/electronics10212689>
- Aitken, M., Toreini, E., Carmichael, P., Coopamootoo, K., Elliott, K., & van Moorsel, A. (2020). Establishing a Social Licence for Financial Technology: Reflections on the Role of the Private Sector in Pursuing Ethical Data Practices. *Big Data & Society*, 7. <https://doi.org/10.1177/2053951720908892>
- Baddeley, A. (1981). The Concept of Working Memory: A View of Its Current State and Probable Future Development. *Cognition*, 10, 17-23. [https://doi.org/10.1016/0010-0277\(81\)90020-2](https://doi.org/10.1016/0010-0277(81)90020-2)
- Cooper, S. J. (2005). Donald O. Hebb's Synapse and Learning Rule: A History and Commentary. *Neuroscience & Biobehavioral Reviews*, 28, 851-874. <https://doi.org/10.1016/j.neubiorev.2004.09.009>
- Danek, A. H., & Salvi, C. (2020). Moment of Truth: Why Aha! Experiences Are Correct. *The Journal of Creative Behavior*, 54, 484-486. <https://doi.org/10.1002/jocb.380>
- Gugerty, L. (2006). Newell and Simon's Logic Theorist: Historical Background and Impact on Cognitive Modeling. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50, 880-884. <https://doi.org/10.1177/154193120605000904>
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex. *The Journal of Physiology*, 160, 106-154. <https://doi.org/10.1113/jphysiol.1962.sp006837>
- Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R. et al. (2004). Neural Activity When People Solve Verbal Problems with Insight. *PLoS Biology*, 2, e97. <https://doi.org/10.1371/journal.pbio.0020097>
- Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A. et al. (2017). Overcoming Catastrophic Forgetting in Neural Networks. *Proceedings of the National Academy of Sciences*, 114, 3521-3526. <https://doi.org/10.1073/pnas.1611835114>
- Luo, J., & Knoblich, G. (2007). Studying Insight Problem Solving with Neuroscientific Methods. *Methods*, 42, 77-86. <https://doi.org/10.1016/j.ymeth.2006.12.005>
- Miller, G. A. (2003). The Cognitive Revolution: A Historical Perspective. *Trends in Cognitive Sciences*, 7, 141-144. [https://doi.org/10.1016/s1364-6613\(03\)00029-9](https://doi.org/10.1016/s1364-6613(03)00029-9)
- Muzaffar, S., & Afshari, A. (2019). Short-Term Load Forecasts Using LSTM Networks. *Energy Procedia*, 158, 2922-2927. <https://doi.org/10.1016/j.egypro.2019.01.952>
- Ohlsson, S. (2011). *Deep Learning: How the Mind Overrides Experience*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511780295>
- Sabharwal, A., & Selman, B. (2011). S. Russell, P. Norvig, Artificial Intelligence: A Modern Approach, Third Edition. *Artificial Intelligence*, 175, 935-937. <https://doi.org/10.1016/j.artint.2011.01.005>
- Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D., & Lillicrap, T. P. (2016). One-Shot Learning with Memory-Augmented Neural Networks. *arXiv:1605.06065*. <https://doi.org/10.48550/arXiv.1605.06065>
- Scholl, B. J., & Nakayama, K. (2004). Illusory Causal Crescents: Misperceived Spatial Relations Due to Perceived Causality. *Perception*, 33, 455-469. <https://doi.org/10.1068/p5172>
- Scolari, M., Seidl-Rathkopf, K. N., & Kastner, S. (2015). Functions of the Human Frontoparietal Attention Network: Evidence from Neuroimaging. *Current Opinion in Behavioral Sciences*, 1, 32-39. <https://doi.org/10.1016/j.cobeha.2014.08.003>