

# 连续学习常见认知任务

## 一、课题背景

人工智能在很多领域上处理特定任务的能力已经达到了人的平均水平，甚至远超人类。通常，衡量人工智能是否成功的一大准则就是判断其模仿人类学习的能力。在特定任务上，机器被给予真实世界或仿真模型中的大量数据用于训练以完成特定任务。但这一过程仅仅关注最终结果，而忽视了学习过程，因此也就不具备人类学习中一大重要特征，即能够灵活地切换需要完成的任务，又可以在训练过程中连续地积累知识和经验。世界是在发展和不断变化的，如果不能具备适应的能力，就很难被称为真正地拥有智能。然而，这样的强鲁棒性和当下主流机器学习（Machine Learning, ML）算法相矛盾，统计机器学习方法大多依赖于独立同分布数据假设，且需要预处理和筛选过、大量、均质化的数据样本进行学习。当数据发生变化或是样本空间增大时，ML 算法常常对新的任务无能为力，或是习得新任务后在先前学习的任务上表现不佳，这也被称为灾难性遗忘（Catastrophic Forgetting）。

为了应对非静态环境下连续出现的任务序列，研究者提出了被称为连续学习（Continual learning）或增量学习（Incremental learning）的方法范式。连续学习帮助模型持续地积累知识，避免灾难性遗忘的发生，使得模型在面临新的知识时无需从头开始训练。当学习的任务相互关联时，当前任务的学习可以帮助模型在每个后续任务上取得更好的性能，或令模型在以前的任务上表现更好，这两种现象被分别称为前向迁移和后向迁移。在本课题中，利用感兴趣的连续学习方法训练模型，比较采用连续学习方法与否的模型表现差异，并分析结果。若能对于现有连续学习方法进行改进，并能借鉴认知神经科学原理、现象的，将视观点的新颖性和深入程度获得额外加分。要求提交完整实现代码，该连续学习方法是否有受到生物智能的启发？如果有，介绍其中涉及的认知神经科学机制。思考人工智能模型连续学习方法和人类认知的异同。

## 二、数据

采用模拟生成的认知任务作为训练集。NeuroGym 是基于 OpenAI Gym 开发的神经科学任务开源 Python 工具箱，提供心理学和认知科学常见的行为范式用于人工智能模型的训练[1]。任务说明参考 [Environments — neurogym documentation](#)。（也可采用其他文献提出的认知任务模拟方式，如 20-Cog-tasks[2] 等。）

## 三、课题内容

(1) 选择合适模型同时完成不少于 10 种认知任务，并给出模型的任务表现。评价指标应至少包含 F1 分数和 MSE 损失中的一种。

(2) 采用不少于 2 种连续学习方法，分别完成认知任务的训练，记录模型在每个任务完成训练时的任务评分。在所有任务完成训练后，对比每一种连续学习训练方式和一般训练方式的任务表现差异。重点关注连续学习是否能有效避免灾难性遗忘，甚至产生前向或后向知识迁移等学习优势。

(3) 在提交的报告中介绍连续学习的认知任务来源于何种认知实验，思考采用的连续学习方法和生物智能体学习方式之间的异同。连续学习后模型是否会更接近生物智能体？若能给出实验证据，例如激活表征、行为表现等，将额外加分。

(4)（可选）改进现有连续学习方法，以提升多任务的连续学习性能。

(5)（可选）考虑到 NEUROGYM 等工具仅提供了认知任务的简单模拟，其仿真实验过程与被试实际完成任务的过程存在诸多差异。改进先前用于训练模型的认知任务，使其更加接近现实世界的实验环境，并分析这种任务仿真形式的改变对模型性能的影响。

## 参考文献

- [1] M. Molano-Mazón *et al.*, “NeuroGym: An open resource for developing and sharing neuroscience tasks,” Feb. 2022, doi: 10.31234/osf.io/aqc9n.
- [2] G. R. Yang, M. R. Joglekar, H. F. Song, W. T. Newsome, and X.-J. Wang, “Task representations

- in neural networks trained to perform many cognitive tasks,” *Nat Neurosci*, vol. 22, no. 2, pp. 297–306, Feb. 2019, doi: 10.1038/s41593-018-0310-2.
- [3] G. M. van de Ven, T. Tuytelaars, and A. S. Tolias, “Three types of incremental learning,” *Nature Machine Intelligence*, vol. 4, no. 12, Art. no. 12, Dec. 2022, doi: [10.1038/s42256-022-00568-3](https://doi.org/10.1038/s42256-022-00568-3).
- [4] R. Saxena, J. L. Shobe, and B. L. McNaughton, “Learning in deep neural networks and brains with similarity-weighted interleaved learning,” *Proc. Natl. Acad. Sci.*, vol. 119, no. 27, p. e2115229119, Jul. 2022, doi: [10.1073/pnas.2115229119](https://doi.org/10.1073/pnas.2115229119).
- [5] M. De Lange *et al.*, “A Continual Learning Survey: Defying Forgetting in Classification Tasks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3366–3385, Jul. 2022, doi: [10.1109/TPAMI.2021.3057446](https://doi.org/10.1109/TPAMI.2021.3057446).
- [6] B. Ehret, C. Henning, M. Cervera, A. Meulemans, J. V. Oswald, and B. F. Grewe, “Continual learning in recurrent neural networks,” presented at the International Conference on Learning Representations, Jan. 2021. Available: <https://openreview.net/forum?id=8xeBUgD8u9>
- [7] G. M. van de Ven, H. T. Siegelmann, and A. S. Tolias, “Brain-inspired replay for continual learning with artificial neural networks,” *Nat. Commun.*, vol. 11, no. 1, p. 4069, Aug. 2020, doi: [10.1038/s41467-020-17866-2](https://doi.org/10.1038/s41467-020-17866-2).
- [8] R. Hadsell, D. Rao, A. A. Rusu, and R. Pascanu, “Embracing Change: Continual Learning in Deep Neural Networks,” *Trends in Cognitive Sciences*, vol. 24, no. 12, pp. 1028–1040, Dec. 2020, doi: [10.1016/j.tics.2020.09.004](https://doi.org/10.1016/j.tics.2020.09.004).
- [9] 朱飞, 张煦尧, and 刘成林, “类别增量学习研究进展和性能评价,” *自动化学报*, vol. 49, no. 3, pp. 635–660, Mar. 2023.
- [10] T. Zhang, X. Cheng, S. Jia, C. T. Li, M. Poo, and B. Xu, “A brain-inspired algorithm that mitigates catastrophic forgetting of artificial and spiking neural networks with low computational cost,” *Sci. Adv.*, vol. 9, no. 34, p. eadi2947, Aug. 2023, doi: [10.1126/sciadv.adj2947](https://doi.org/10.1126/sciadv.adj2947).
- [11] G. Zeng, Y. Chen, B. Cui, and S. Yu, “Continual learning of context-dependent processing in neural networks,” *Nat Mach Intell*, vol. 1, no. 8, pp. 364–372, Aug. 2019, doi: 10.1038/s42256-019-0080-x.