A Predictive Processing Implementation of the Common Model of Cognition

Alexander Ororbia (ago@cs.rit.edu)
Department of Computer Science, Rochester Institute of Technology, Rochester, NY 14623

M. Alex Kelly (AlexKelly@cunet.carleton.ca)
Department of Computer Science, Bucknell University, Lewisburg, PA, USA
Department of Cognitive Science, Carleton University, Ottawa, ON, Canada

Keywords: cognitive architecture; predictive processing; holographic memory; continual learning

While modern machine learning techniques based on deep artificial neural networks (ANNs) have an impressive ability to process data to uncover patterns, they do not typically model high-level cognition or more than a single task. If an ANN is trained on a series of tasks, catastrophic interference occurs, with each new task causing the ANN to forget all previously learned tasks (McCloskey & Cohen, 1989). Conversely, symbolic cognitive architectures can capture the complexities of high-level cognition but scale poorly to the naturalistic, non-symbolic data of sensory perception (e.g., images) or to big datasets necessary for modelling learning over a lifetime (e.g., corpora with hundreds of millions of words). Is it possible to provide a theory that bridges ANNs and symbolic models, a reduction of the symbolic to the neural, while retaining the strengths and capabilities of each?

We propose a cognitive architecture that is built on two biologically plausible, neural models: neural generative coding (NGC; Ororbia, Mali, Giles, & Kifer, 2020) and holographic memory (Kelly, Arora, West, & Reitter, 2020). By combining the two, we create a model of cognition that has the power of modern machine learning techniques while retaining long-term memory, single-trial learning, transfer-learning, and other cognitive capacities associated with high-level cognition. Our intent is to advance towards a cognitive architecture capable of capturing human performance at all scales of learning, from the half-hour lab experiment to skills acquired gradually over a lifetime.

Since Newell (1973) first argued that good empirical work and piecemeal theoretical work are insufficient to understand the mind, researchers in cognitive science have sought to develop functional, testable theories of cognition as a whole. Cognitive architectures serve as both unified theories of cognition and as computational frameworks for implementing models of specific experimental tasks. Forty years of research has developed hundreds of cognitive architectures with strong commonalities to each other (Kotseruba & Tsotsos, 2018) suggesting an emerging consensus on the basic principles of cognition, on the basis of which Laird et al. (2017) propose a *Common Model of Cognition*, a high-level theory of the modules of the mind and how they interact (see Fig. 1).

The Common Model of Cognition consists of perceptual

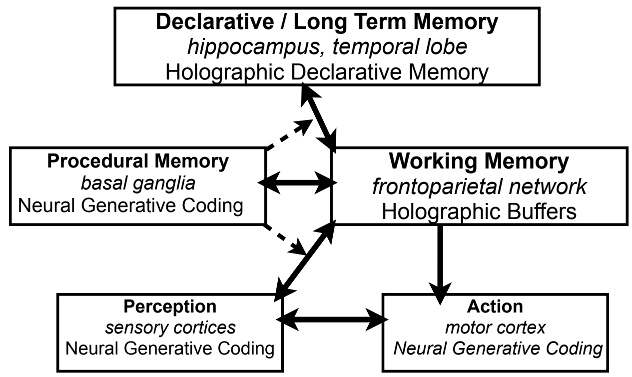

Figure 1: Common Model of Cognition (Laird et al., 2017), associated brain areas (Stocco et al., 2021), and our approach to modelling each module. Solid arrows are data passing. Dashed arrows indicate modulation of a data passing path.

and motor modules that interact with the agent's environment, working memory to hold the active data in the agent's mind, a declarative memory that holds the agent's world knowledge, and a procedural memory that controls information and evaluates possible actions. An evaluation of fMRI data from 200 participants across tasks found correlations in patterns of activity across brain areas consistent with the Common Model of Cognition's modules (Stocco et al., 2021).

Proposed Architecture

Neural Generative Coding (NGC) is a scalable instantiation of predictive processing brain theory (Clark, 2015) yielding an efficient, robust form of predict-then-correct learning.

Neural Generative Sensory Cortices use NGC for processing a specific modality of data. In Ororbia et al. (2020), we show that NGC learns a good density estimator of data (from which new samples can be sampled or "fantasized"), in conjunction with desired target functionality (e.g., classification, regression), in not only the cases of static input but also in cases of time-varying data series.

Neural Generative Motor Cortex In Ororbia and Mali (2021), we generalize NGC to the case of action-driven tasks, i.e., active NGC (ANGC), common in reinforcement learning (RL), providing evidence that NGC can be used to build a coupled generative model and controller system that solves RL problems, particularly those when the reward signal is sparse or non-existent. ANGC will serve as the motor cortex.

Neural Generative Basal Ganglia In Ororbia, Mali, Kifer, and Giles (2019), we model the functionality of the basal ganglia in suppressing/inhibiting neural activity for the purpose of action selection and task switching (Cameron, Watanabe, Pari, & Munoz, 2010), a behavior we argue is critical in facilitating effective continual learning without catastrophic interference. This task selection model, which learns through competitive Hebbian learning, will serve as the basis for part of the basal ganglia in our architecture, acting to coordinate the exchange of information between the working memory and the sensory and long-term memory cortices.

Holographic memory (Plate, 1995) is a formalism for capturing the capacity for humans to learn and recall arbitrarily complex associations between stimuli in the environment. Holographic memory is immune to the catastrophic interference typical of more conventional ANNs (Mannering & Jones, 2021), allowing it to be used to construct models that handle multiple, unrelated tasks (Cheung, Terekhov, Chen, Agrawal, & Olshausen, 2019).

Working Memory Each buffer in working memory is a holographic vector. Holographic memory vectors have an established ability to account for memory phenomena such as serial and free recall of lists (Franklin & Mewhort, 2015).

Declarative Memory is the composition of many individual holographic vectors (each representing a distinct concept). Our model accounts for human performance in recall, probability judgement, decision-making (Kelly, Arora, et al., 2020), and learning the meaning and part-of-speech of words (Kelly, Ghafurian, West, & Reitter, 2020).

Conclusions and Future Research

Humans are capable of continual learning, deep expertise, single-trial learning and agile adaptation to dynamic environments, and transfer learning across multiple tasks. Conventional ANNs struggle to replicate these abilities. Solving the problem of lifelong learning will aid us both in understanding the human mind and in the development of intelligent agents that are better able to generalize to real world environments. Our proposed implementation of the Common Model of Cognition is composed of neuro-cognitively plausible components, i.e., holographic memory, predictive processing circuits, and competitive learning. A promising research direction is the application of our architecture to where the challenge of catastrophic interference is most prevalent: reinforcement learning across lengthy, diverse streams of tasks where knowledge retrieval, transfer, and composition are absolutely critical.

References

Cameron, I. G., Watanabe, M., Pari, G., & Munoz, D. P. (2010). Executive impairment in parkinson's disease: response automaticity and task switching. *Neuropsychologia*, *48*(7), 1948–1957.

Cheung, B., Terekhov, A., Chen, Y., Agrawal, P., & Olshausen, B. (2019). Superposition of many models into one. In *Adv Neural Inf Process Syst 32* (pp. 10868–10877).

Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind.* Oxford University Press.

Franklin, D. R. J., & Mewhort, D. J. K. (2015). Memory as a hologram: An analysis of learning and recall. *Canadian Journal of Experimental Psychology*, *69*, 115-135.

Kelly, M. A., Arora, N., West, R. L., & Reitter, D. (2020). Holographic declarative memory: Distributional semantics as the architecture of memory. *Cognitive Science*, *44*(11), e12904. doi: 10.1111/cogs.12904

Kelly, M. A., Ghafurian, M., West, R. L., & Reitter, D. (2020). Indirect associations in learning semantic and syntactic lexical relationships. *Journal of Memory and Language*, *115*, 104153. doi: 10.1016/j.jml.2020.104153

Kotseruba, I., & Tsotsos, J. K. (2018). 40 years of cognitive architectures: Core cognitive abilities and practical applications. *Artificial Intelligence Review*.

Laird, J. E., Lebiere, C., & Rosenbloom, P. S. (2017). A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *AI Magazine*, *38*(4), 13–26. doi: 10.1609/aimag.v38i4.2744

Mannering, W. M., & Jones, M. N. (2021). Catastrophic interference in predictive neural network models of distributional semantics. *Comput Brain Behav*, *4*, 18–33.

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. *The psychology of learning and motivation*, *24*(109), 92.

Newell, A. (1973). You can't play 20 questions with nature and win. In W. G. Chase (Ed.), *Visual information processing* (p. 283-308). New York: Academic Press.

Ororbia, A. G., & Mali, A. (2021). Adapting to dynamic environments with active neural generative coding. doi: 10.13140/RG.2.2.14273.48485/1

Ororbia, A. G., Mali, A., Giles, C. L., & Kifer, D. (2020). Continual learning of recurrent neural networks by locally aligning distributed representations. *IEEE Transactions on Neural Networks and Learning Systems*.

Ororbia, A. G., Mali, A., Kifer, D., & Giles, C. L. (2019). Lifelong neural predictive coding: Learning cumulatively online without forgetting. *arXiv:1905.10696*.

Plate, T. A. (1995). Holographic reduced representations. *IEEE Transactions on Neural networks*, *6*(3), 623–641.

Stocco, A., Sibert, C., Steine-Hanson, Z., Koh, N., Laird, J. E., Lebiere, C. J., & Rosenbloom, P. (2021). Analysis of the human connectome data supports the notion of a "common model of cognition" for human and human-like intelligence across domains. *NeuroImage*, *235*, 118035.