Toward Human-Like Representation Learning for Cognitive Architectures

Steven Jones, Peter Lindes
Center for Integrated Cognition
24 Frank Lloyd Wright Dr
Ann Arbor, Michigan 48105 USA
{steven.jones,peter.lindes}@cic.iqmri.org

Abstract

Human-like learning includes an ability to learn concepts from a stream of embodiment sensor data. Echoing previous thoughts such as those from Barsalou that cognition and perception share a common representation system, we suggest an addendum to the common model of cognition. This addendum poses a simultaneous semantic memory and perception learning that bypasses working memory, and that uses parallel processing to learn concepts apart from deliberate reasoning. The goal is to provide a general outline for how to extend a class of cognitive architectures to implement a more human-like interface between cognition and embodiment of an agent, where a critical aspect of that interface is that it is dynamic because of learning.

The Common Model of Cognition (CMC) describes a set of cognitive architectures (CAs) that implement many cognitive capabilities, including many human-like learning capabilities (Laird, Lebiere, and Rosenbloom 2017). Many of these architectures focus on symbolic processing with connections over fixed interfaces to separate perceptual modules. The symbolic learning in these architectures can be considered Machine Learning (ML), but is very different from Deep Learning (DL) where a large neural network is trained from big data offline.[1]

There are key features that distinguish the DL approach from the CA approach. DL typically involves 1) big data for training, 2) massive parallelism, and 3) offline training to produce a static structure for online use. Human-like perceptual representation learning is somewhere between CA and DL approaches. The big data input for humans is like a streaming concatenated vector of vision, sound, and proprioceptive signals processed with massive parallelism. Unlike much current CA and DL practice, in humans perceptual representation learning occurs online.

We refer to one agent design pattern in Soar (Laird 2012) as an example of where CAs have fallen short of implementing human-like representation learning. In the past, some Soar agents have been developed that use a combination of externally-developed perceptual modules (e.g. DL) and a static interface with Soar's control representations. For example, we've used external perception to derive bounding boxes and object properties, then used those bounding boxes within Soar's systems for spatial and visual reasoning (Kirk, Mininger, and Laird 2016). This integration pattern can lead to successful mobile robot agents capable of interacting with real objects, but the resulting capabilities fail to be completely human-like in part because of the static opaque interface between the cognition and embodiment of the agent.

Human perception transforms signals that enter through sensory neurons into symbolic representations that symbolic reasoning can work with. This transformation takes place in a hierarchy. For speech, for example, audio signals are processed temporally and spectrally to produce a hierarchy of phonemes, syllables, and words (Kuhl 2000). A similar hierarchy is involved in visual processing (Kruger et al. 2012). In humans the signal to symbol transformation is learned incrementally from many experiences over a lifetime. During operation, real-time human perception and memory integrates modality-specific sensor-level representations in a nested hierarchy with abstract situation model representations (Baldassano et al. 2017). We desire learning capable of inferring a similar breadth of representation abstraction and timescale, that scales to human-level sensory data streams.

> *"On this view, a common representational system underlies perception and cognition, not independent systems."* (Barsalou 1999)

> *"Coming to grips with these phenomena of perceptual decisions will force a significant addition to the Soar architecture."* (Newell 1990)

To specify what might be missing from the CMC view that could generally constitute a more human-like learning integrated with CAs, we refer to results suggesting additional architectural connections not included within the CMC. Consider that Perception and Action representations may feed directly to Long-Term Memory (Hake, Sibert, and Stocco 2022). Inspired by that work, we speculate that a human-like ML infers a hierarchy of symbolic representations with subsymbolic metadata spanning from "low-level" perception to declarative causal models of environment and action dynamics using parallel processing and without requiring interaction with Working Memory. We propose this kind of inference (e.g. like with incremental PCFG induc-

[1]However, some work has been done on more integrated approaches such as with Sigma (Rosenbloom, Demskia, and Ustuna 2017) and Spaun (Stewart, Choo, and Eliasmith 2012).

tion by Dechter (2018), but parallelized) as the target of a CA perceptual representation learning.

We expect the benefits of such learning to be: 1) Many embodiment models may be simple and can be learned through inference mechanisms that fall short of the power of extended procedural memory-driven reasoning. These mechanisms may be more easily parallelizable while also adequate for simple representation learning. This divorcing of some representation induction from deliberate reasoning reduces demands on reasoning. 2) Inferred representations that are more directly referent to embodiment signals and in the same representational space as cognition provides cognition more control over its interface with the embodiment than a static opaque externally-derived interface.

References

Baldassano, C.; Chen, J.; Zadbood, A.; Pillow, J. W.; Hasson, U.; and Norman, K. A. 2017. Discovering event structure in continuous narrative perception and memory. *Neuron*, 95(3): 709–721.

Barsalou, L. W. 1999. Perceptual symbol systems. *Behavioral and brain sciences*, 22(4): 577–660.

Dechter, E. 2018. *Using the language of thought*. Ph.D. thesis, Massachusetts Institute of Technology.

Hake, H. S.; Sibert, C.; and Stocco, A. 2022. Inferring a Cognitive Architecture from Multitask Neuroimaging Data: A Data-Driven Test of the Common Model of Cognition Using Granger Causality. *Topics in Cognitive Science*, 14(4): 845–859.

Kirk, J.; Mininger, A.; and Laird, J. 2016. Learning task goals interactively with visual demonstrations. *Biologically Inspired Cognitive Architectures*, 18: 1–8.

Kruger, N.; Janssen, P.; Kalkan, S.; Lappe, M.; Leonardis, A.; Piater, J.; Rodriguez-Sanchez, A. J.; and Wiskott, L. 2012. Deep hierarchies in the primate visual cortex: What can we learn for computer vision? *IEEE transactions on pattern analysis and machine intelligence*, 35(8): 1847–1871.

Kuhl, P. K. 2000. A new view of language acquisition. *Prodeedings of the National Academy of Sciences*, 97(22): 11850–11857.

Laird, J. E. 2012. *The Soar cognitive architecture*. MIT press.

Laird, J. E.; Lebiere, C.; and 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.

Newell, A. 1990. *Unified theories of cognition*. Harvard University Press.

Rosenbloom, P. S.; Demskia, A.; and Ustuna, V. 2017. Toward a Neural-Symbolic Sigma: Introducing Neural Network Learning. In *15th International Conference on Cognitive Modeling*, 73.

Stewart, T.; Choo, F.-X.; and Eliasmith, C. 2012. Spaun: A perception-cognition-action model using spiking neurons. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 34.