Is Deep Learning the Answer for Understanding Human Cognitive Dynamics?

John P. Spencer (j.spencer@uea.ac.uk)
School of Psychology, University of East Anglia
Norwich NR4 7TJ UK

Keywords: natural language processing, machine learning, artificial intelligence, cognitive neuroscience, neural networks

A Revolution in the Making?

Deep Learning is a neural network approach where a network of multiple layers is trained to process complex data. Applications for deep learning include recognizing complex patterns in pictures, text, and sounds to, in some cases, produce insights into how human process information. Deep learning has garnered widescale interest, fostered by advances in, for instance, natural language processing and the release of innovative tools like Chat GPT. But does deep learning have implications for theory in the cognitive sciences; that is, is deep learning the answer for understanding human cognitive dynamics?

Here's what ChatGPT had to say: "Deep learning has shown great promise in certain aspects of mimicking human cognitive processes. However, human cognitive dynamics are incredibly complex and involve a wide range of processes, including perception, memory, reasoning, decision-making, emotions, and more. While deep learning models can be trained to perform specific tasks that involve aspects of cognition, they often lack the holistic understanding and flexibility that characterize human cognition." What do the experts have to say? In this symposium, our distinguished speakers will address the strengths and weaknesses of using deep learning to advance our theoretical understanding of human cognitive dynamics.

Presenter Qualifications

Moderator: John P. Spencer John Spencer is a Professor of Psychology at the University of East Anglia in Norwich, UK. He received a Ph.D. from Indiana University in 1998. His research examines the development of word learning, working memory, attention, and executive function using brain imaging and Dynamic Field Theory.

Talk 1: Brenden Lake Brenden Lake is an Assistant Professor of Psychology and Data Science at New York University. He received his PhD in 2014 from the Massachusetts Institute of Technology. Prof Lake's research seeks the ingredients of intelligence. He uses advances in machine intelligence to better understand human intelligence, and uses insights from human intelligence to develop more fruitful kinds of machine intelligence.

Talk 2: Raul Grieben & Gregor Schöner Raul Grieben is a PhD student in the lab of Prof. Dr. Gregor Schöner at the Ruhr University, Bochum. Prof. Schöner received his PhD in 1985 from the University of Stuttgart. Prof. Schöner's group examines how embodied and situated nervous systems develop cognition. His group has pioneered the use Dynamic Field Theory to systematically build an account of action, perception, and embodied cognition that have been tested using robotics and simulations of human data.

Talk 3: Mariya Toneva Mariya Toneva is a faculty member at the Max Planck Institute for Software Systems. She received her PhD in 2021 from Carnegie Mellon University. Her research is at the intersection of Machine Learning, Natural Language Processing, and Neuroscience, with a focus on building models of language processing in the brain that can also improve natural language processing systems.

Talk 4: Gina Kuperberg Gina R Kuperberg, MD PhD, is the Dennett Stibel Professor of Cognitive Science at Tufts University, and a Principal Investigator in the Psychiatry Neuroscience Program at Massachusetts General Hospital. Her lab uses multimodal neuroimaging techniques (fMRI, MEG and EEG evoked and multivariate responses), together with modeling, to understand when, where and how the human brain builds meaning from language, and how these mechanisms break down in neuropsychiatric disorders.

Talk 1: Neural network modeling through the eyes and ears of a child

Young children have sophisticated, dynamic representations of their visual and linguistic environment. Where do these representations come from? How much knowledge arises through generic learning mechanisms applied to sensory data, and how much requires more substantive (possibly innate) inductive biases? Using deep learning, we examine these questions by training relatively generic models solely on longitudinal data collected from a single child (Sullivan et al., 2020), consisting of egocentric video and audio streams. Our principal findings are as follows: 1) Based on visual only training, neural networks can acquire high-level visual features that are broadly useful across categorization and segmentation tasks. 2) Based on language only training, networks can acquire meaningful clusters of words and sentence-level syntactic sensitivity. 3) Based on paired visual and language training, networks can acquire word-referent mappings from tens of noisy examples and align their multi-modal conceptual systems. Taken together, our results show how sophisticated visual and linguistic representations can arise through data-driven learning applied to one child's first-person experience. We'll also discuss the ways that the learning dynamics of these models both match and mismatch the dynamics of human cognitive development.

Talk 2: Bridging Dynamic Field Theory and Deep Neural Networks – A model of guided

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

visual attention in natural environments

Our world has a well-structured environment that shapes our expectations and influences how we interact with objects around us. However, most studies in visual cognition use artificial visual scenes and simplified stimuli, which limits our understanding of how humans visually explore and search for real-world objects in natural scenes. Attention is also crucial in modern machine-learning architectures; however, there is significant divergence between the comprehension of attention in deep learning and in psychology. While deep neural networks (DNNs) outperform humans in visual object categorization, they fail to replicate human visual search performance. Why? Visual search and voluntary attention require a recurrent feedback loop and stable memory representations for goal-oriented interaction with the environment.

Dynamic Field Theory (DFT) provides the necessary neural processes to understand naturalistic visual search, including autonomous processes, sequence generation, and working memory. We present the latest version of our neural dynamic process model that combines the strengths of DFT and DNNs to explain: 1) how a mapping from the distributed representation of a convolutional neural network (CNN) to the localist representation of a dynamic neural field may be learned; 2) how guidance templates learned from visual input enable categorical guided visual search; 3) how guidance templates in working memory may be adapted to different tasks; 4) how scene grammar can emerge as a special case of scene guidance in natural scenes; 5) how functional visual fields influence search.

Our model's biggest strength, besides its neural plausibility, is its operation in a closed behavioral loop, which sets it apart from an end-to-end learned CNN. Stable memory representations enable goal-oriented actions, while adaptive recurrent top-down feedback allows flexible switching between modes without specific algorithms.

Talk 3: Large language models are useful model organisms of language processing in the human brain

When reading the sentence "The trophy doesn't fit into the brown suitcase because it's too big", we understand the meaning of this sentence despite the ambiguous pronoun "it", which may refer to either the trophy or the suitcase. How does the brain process this sentence and attribute real-world meaning to it? To address this, there are some fundamental preliminary questions to answer about what information is processed where and when in the brain to understand how this information is aggregated across different locations and time points. Using neuroimaging devices that record human brain activity during language processing, neuroscientists have made progress towards answering the what, where, and when questions. However, how the meaning of words is aggregated together by the brain remains elusive.

Meanwhile, recent advances in large language models (LLMs) have created computational systems that aggregate the meaning of words in specific ways to perform a specific linguistic task, such as predicting the upcoming word in a sentence. However, it is not clear whether these computational systems truly understand the meaning of a sentence, and whether the "how" of an LLM is the same as the "how" of the brain. In this talk, I present evidence that neurolinguistics can benefit from using LLMs as model organisms for how information is aggregated during language comprehension in the human brain, despite LLM's differences from the human brain.

Talk 4: "Black-box tools" and "Brain-inspired modeling" provide complementary insights into the representations and neural dynamics underlying language processing

Box's famous assertion that "All models are wrong, but some are useful" has become a catchphrase among cognitive scientists. However, for our models to be useful, we must tailor them to our questions of interest. To do this, my lab is using two distinct modeling strategies to ask when, where, and how the brain infers meaning from language.

First, we employ a diverse set of Natural Language Processing (NLP) models as theoretically motivated tools to probe distinct representations across the linguistic hierarchy during natural language processing. We integrate these models with time-sensitive neuroimaging techniques (EEG/MEG) to determine where and when the brain builds these representations during typical language comprehension. We are also leveraging this approach to characterize the disorganized speech patterns produced by some people with schizophrenia during natural language production ("positive thought disorder").

Most "black box" NLP architectures, however, have little in common with the neurobiological and cognitive architecture of the human brain. Therefore, to understand (a) how information interacts across the linguistic/cortical hierarchy, and (b) how this gives rise to the dynamics of neural activity evoked by each word during real-time processing, we built a biologically and cognitively plausible, predictive coding model of lexico-semantic processing. This model is small-scale, with pre-specified representations at each level of its hierarchy, meaning that its dynamics are highly interpretable. Our simulations show that predictive coding (a) explains how information is propagated up and down the linguistic hierarchy, (b) predicts where these effects localize across the cortical hierarchy, (c) captures complex interactions between top-down contextual effects and bottom-up lexical effects on both neural activity and behavior, and (d) provides a natural, intuitive, and biologically plausible explanation for the time-course of both univariate and multivariate neural activity.

Together, our "NLP models as tools" and "brain-inspired modeling" approaches are highly complementary, with each providing unique insights into both healthy and atypical language processing.