Published in: Current Opinion in Neurobiology (2019). 55:iii-iv

Editorial overview: Machine learning, big data, and neuroscience

Jonathan W. Pillow & Maneesh Sahani

May 31, 2019

Abstract

Editorial introduction to a special issue of Current Opinion in Neurobiology on "Machine learning, big data, and neuroscience".

The articles in this issue explore the growing use and uses of machine-learning methods in neuroscience. The field of machine learning is a relatively recent one, developing in the 1980s through a convergence of earlier ideas in computer science, statistics, and connectionist psychology. At its heart, machine learning seeks to construct and to characterise algorithms that adapt to their input. Machine-learning algorithms may perform different sorts of task—they may make predictions, infer underlying structure or causes in input, or select actions to achieve a defined goal. But in all cases, the explicit rules of prediction, inference or action are not programmed by the designer. Instead, the algorithm is designed to discover these rules from training examples, by trial and error, or by uncovering statistically reliable structure in the inputs alone.

In the past decade or so clever new learning algorithms, combined with the availability of large training data sets and the relentless advance of computing horsepower, have brought machine learning to the forefront of many applications; from functional stereochemistry to image recognition and processing, to playing the game of Go. Indeed, machine learning methods have fueled a renaissance in artificial intelligence, which now promises to bring to fruition many long-standing technological aspirations.

It should be no surprise that the connections between neuroscience and machine learning run deep. Early connectionism—from Rosenblatt's (1958) perceptron onwards—was itself inspired by the distributed nature of neural systems; and a sense that many parts of the brain (particularly circuits in the neocortex and the cerebellum) were essentially *tabulae rasae*, to be shaped primarily by interactions with the sensory environment. The architectures explored by connectionism led directly to the distributed "neural network" function approximators used extensively in machine learning; including the deep convolutional structures which have revolutionised image processing (Fukushima, 1980).

Thus neuroscience has provided architectural metaphors that have proven important in parts of machine learning. Conversely, machine learning has provided functional metaphors, which can guide hypotheses about and explorations of neural function. Perhaps the most remarkable convergence of this sort has been between the subfield of reinforcement learning (RL), which studies learning in agents whose actions shape their future environment as well as their success, and the neuroscience of the basal ganglia. Most famously, Schultz et al. (1997) suggested that the phasic activity of subcortical dopaminergic neurons corresponded to a reward prediction error signal of exactly the sort required in the fundamental RL algorithm known as temporal-difference learning (Sutton, 1988). Convergence between the results of RL and descriptions of both neural activity and animal behaviour have grown to encompass a wide range of functions. Schulz and Gershman (2019) and Cortese et al. (2019) explore some contemporary themes in this collection. Perhaps most remarkable is that, beyond normal function, RL has also begun to provide insight into the possible computational basis of many psychiatric diseases (Montague et al., 2012), a theme that is explored further by Radulescu and Niv (2019).

A functional metaphor from machine learning that comes close second to RL in its influence within neuroscience, is that of "representational learning": broadly, the algorithmic adaptation of a distributed encoding of sensory input. In one approach, aligned with *unsupervised machine learning*, this adaptation is driven by statistical regularity within natural inputs — exemplified by the influential results of Olshausen and Field (1996). Recent studies have elaborated extensively on the basic model of natural stimuli, with Sanchez-Giraldo et al. (2019) discussing one such extension here. A significant recent line of work has also investigated the potential for artificial networks trained using *supervision* on naturalistic tasks such as object classification to yield representations of neural relevance; and this approach is explored by both Kell and McDermott (2019) and Barrett et al. (2019) in this issue. Beyond such correspondence with early neural representations, an exciting contemporary frontier is opened by the potential for machine learning to provide a functional metaphor within which to understand higher-level cognitive representation—a theme explored by Spicer and Sanborn (2019) and Yildirim et al. (2019) here.

Correspondences in representation are drawn between the outcomes of machine learning algorithm and the pattern of neural responses observed in experiments. Such correspondences raise a natural question: might it be that the algorithmic processes that arrive at such representations in machine settings are also reflected within learning in the brain? This remains a contentious issue. Perhaps the most suggestive correspondence to the algorithms of supervised learning is the match between single-layer perceptron learning and the Marr-Albus-Ito theory of learning in the cerebellum (Albus, 1989). However correspondence in deeper networks, or in recurrent network structures has been difficult to establish. This issue is explored here by Lillicrap and Santoro (2019).

These forms of convergence with machine learning at the algorithmic and possibly architectural levels are unique to neuroscience—a field that studies biological systems that adapt to information-rich environments and complex goals exactly as machine-learning systems aspire to do. But neuroscience also benefits, in the same way as other fields, from the remarkable power of machine learning methods to process, condense and find patterns within data. The volume and complexity of neuroscience data have burgeoned in a technological revolution that has paralleled that in machine learning, and neither manual inspection, nor simple statistical methods provide the scale or power needed to analyse modern data from connectomic imaging to dense kiloscale eletrophysiology.

The first challenge in the analysis of neural data is to take raw data (e.g. electron-micrographs, optical or electrical measurements) and extract meaningful anatomical or functional signals at the level of neurons or their connections. Machine learning methods play a crucial role in such 'signal-path' processing. Several articles in this volume cover recent advances in methods for processing different kinds of data, ranging from methods for connectomics and anatomy at the cellular and sub-cellular level (Lee et al., 2019; Rolnick and Dyer, 2019; Motta et al., 2019; Vogelstein et al., 2019) to methods for inferring functional connectivity at the whole-brain level (Foti and Fox, 2019). Methods for extracting neural activity from raw signals are addressed for calcium fluorescence measurements (Pnevmatikakis, 2019; Stringer and Pachitariu, 2019), as well as for electrical signals, a problem setting known as spike sorting (Carlson and Carin, 2019).

Image segmentation, spike isolation and similar signal processing algorithms convert raw data in pixels and samples into measurements of neurally meaningful quantities such as neuronal activation or connectivity. But even these variables form rich and varied datasets spanning many neurons with complex relationships. To work out what they say about neural function is likely to require tools of much greater sophistication than averaging or similar aggregation. Here too, then, machine learning methods have a key role to play. Two articles in the current volume focus on the analysis and interpretation of neural population responses: Saxena and Cunningham (2019) articulate a 'neuron population doctrine', arguing that neural populations should be considered a fundamental unit of computation in the brain, while Williamson et al. (2019) focus on methods for connecting large-scale theoretical models with large-scale recordings of neural activity via dimensionality reduction. A second pair of articles focus on the use of encoding and decoding models for interpreting neural datasets (Kriegeskorte and Douglas, 2019), and on the power of such models for moving beyond the limits of specific paradigms and tasks (Varoquaux and Poldrack, 2019).

Finally, machine learning holds enormous promise for the production of medical knowledge and new methodologies for the treatment of brain disfunction and disorder. In this volume, Rao (2019) exlores the potential of modern machine learning to transform brain–computer interfaces from one-way devices that use brain signals to control external prostheses into two-way 'co-processors' that also allow external signals to affect brain states. Cornblath et al. (2019) and Rutledge et al. (2019) focus on applications in the domain of neuropsychiatry, in particular on methods for harnessing large-scale datasets for the design of new models and new therapies for neuropsychiatric illness.

Both neuroscience and machine learning are rapidly developing fields. The points of convergence discussed in this volume provide a snapshot of what has been and will undoubtedly continue to be a rich and meaningful interaction between the two disciplines.

References

Albus, J. S. (1989). The marr and albus theories of the cerebellum-two early models of associative memory. In *Digest of Papers. COMPCON Spring 89. Thirty-Fourth IEEE Computer Society International Conference: Intellectual Leverage*, pages 577–582.

Barrett, D. G., Morcos, A. S., and Macke, J. H. (2019). Analyzing biological and artificial neural networks: challenges with opportunities for synergy? *Current opinion in neurobiology*, 55:55–64.

Carlson, D. and Carin, L. (2019). Continuing progress of spike sorting in the era of big data. *Current opinion in neurobiology*, 55:90–96.

Cornblath, E. J., Lydon-Staley, D. M., and Bassett, D. S. (2019). Harnessing networks and machine learning in neuropsychiatric care. *Current opinion in neurobiology*, 55:32–39.

Cortese, A., De Martino, B., and Kawato, M. (2019). The neural and cognitive architecture for learning from a small sample. *Current Opinion in Neurobiology*, 55:133–141.

Foti, N. J. and Fox, E. B. (2019). Statistical model-based approaches for functional connectivity analysis of neuroimaging data. *Current opinion in neurobiology*, 55:48–54.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202.

Kell, A. J. and McDermott, J. H. (2019). Deep neural network models of sensory systems: windows onto the role of task constraints. *Current opinion in neurobiology*, 55:121–132.

Kriegeskorte, N. and Douglas, P. K. (2019). Interpreting encoding and decoding models. *Current opinion in neurobiology*, 55:167–179.

Lee, K., Turner, N., Macrina, T., Wu, J., Lu, R., and Seung, H. S. (2019). Convolutional nets for reconstructing neural circuits from brain images acquired by serial section electron microscopy. *Current opinion in neurobiology*, 55:188–198.

Lillicrap, T. P. and Santoro, A. (2019). Backpropagation through time and the brain. *Current opinion in neurobiology*, 55:82–89.

Montague, P. R., Dolan, R. J., Friston, K. J., and Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Science*, 16(1):72–80.

Motta, A., Schurr, M., Staffler, B., and Helmstaedter, M. (2019). Big data in nanoscale connectomics, and the greed for training labels. *Current opinion in neurobiology*, 55:180–187.

Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–9.

Pnevmatikakis, E. A. (2019). Analysis pipelines for calcium imaging data. *Current opinion in neurobiology*, 55:15–21.

Radulescu, A. and Niv, Y. (2019). State representation in mental illness. *Current opinion in neurobiology*, 55:160–166.

Rao, R. P. (2019). Towards neural co-processors for the brain: combining decoding and encoding in brain–computer interfaces. *Current opinion in neurobiology*, 55:142–151.

Rolnick, D. and Dyer, E. L. (2019). Generative models and abstractions for large-scale neuroanatomy datasets. *Current opinion in neurobiology*, 55:112–120.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.

Rutledge, R. B., Chekroud, A. M., and Huys, Q. J. (2019). Machine learning and big data in psychiatry: toward clinical applications. *Current opinion in neurobiology*, 55:152–159.

Sanchez-Giraldo, L. G., Laskar, M. N. U., and Schwartz, O. (2019). Normalization and pooling in hierarchical models of natural images. *Current opinion in neurobiology*, 55:65–72.

Saxena, S. and Cunningham, J. P. (2019). Towards the neural population doctrine. *Current opinion in neurobiology*, 55:103–111.

Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599.

Schulz, E. and Gershman, S. J. (2019). The algorithmic architecture of exploration in the human brain. *Current opinion in neurobiology*, 55:7–14.

Spicer, J. and Sanborn, A. N. (2019). What does the mind learn? a comparison of human and machine learning representations. *Current opinion in neurobiology*, 55:97–102.

Stringer, C. and Pachitariu, M. (2019). Computational processing of neural recordings from calcium imaging data. *Current opinion in neurobiology*, 55:22–31.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, 3(1):9–44.

Varoquaux, G. and Poldrack, R. A. (2019). Predictive models avoid excessive reductionism in cognitive neuroimaging. *Current opinion in neurobiology*, 55:1–6.

Vogelstein, J. T., Bridgeford, E. W., Pedigo, B. D., Chung, J., Levin, K., Mensh, B., and Priebe, C. E. (2019). Connectal coding: discovering the structures linking cognitive phenotypes to individual histories. *Current opinion in neurobiology*, 55:199–212.

Williamson, R. C., Doiron, B., Smith, M. A., and Byron, M. Y. (2019). Bridging large-scale neuronal recordings and large-scale network models using dimensionality reduction. *Current opinion in neurobiology*, 55:40–47.

Yildirim, I., Wu, J., Kanwisher, N., and Tenenbaum, J. (2019). An integrative computational architecture for object-driven cortex. *Current opinion in neurobiology*, 55:73–81.