NeuroImage 180 (2018) 117–118

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Outlook on deep neural networks in computational cognitive neuroscience

Brandon M. Turner *

The Ohio State University, United States

Steven Miletić, Birte U. Forstmann

University of Amsterdam, The Netherlands

The contributions of deep neural networks (DNNs), especially the ones appearing in this special issue, clearly demonstrate powerful machinery for understanding the computations that might underlay various processes, especially those observed in vision. Adopting Marr's (1982) levels of analyses, we can think of DNNs as targeting the lowest level of computation: the implementation level. Here, DNNs are thought to mimic neuron-like units that exhibit very complex interactions among one another in order to achieve a sense of representation.

As a measure of success, DNNs are often evaluated on how well they are able to predict the object category of a given stimulus, or more generally, how well the representation produced by the system matches what was presented. To achieve high accuracy, DNNs must be constructed by way of training (e.g., through back propagation) to produce similar computations as the neuronal systems, and as such, they must also be complex. This tendency is often referred to as Bonini's paradox (Bonini, 1963), where in the pursuit of fully understanding a complex system, a tradeoff ensues between the complexity of the model and the accuracy of its computations. The natural endpoint of the model building process is a model that is nearly perfectly accurate, but is also impenetrably complex.

While having complex models is not a problem by itself, the degree to which a model is complex should be justified by empirical data about both brain and cognition. Of course, no one would argue that understanding the brain or cognition is an easy problem, but without a parsimonious model, it will always be difficult to understand what the model provides in terms of knowledge about the system.

Beyond the issue of complexity is the issue of cognition, specifically the modulatory effects that cognition may have on the neural system. It is well known that neural processing is susceptible to attention, task instructions, or other top-down processes. These modulators comprise Marr's computational level, and they can have a direct effect on how representations are used at both the algorithmic and implementation levels. Because of top-down influences from the computational and algorithmic levels, the details of the implementation level can vary from one task to the next. Without an appropriate theory for how the representations used in DNNs should be used across tasks, DNNs can only describe a system within a specific task. To learn a new task, another layer would be needed as well as a more elaborated training period. Are there other ways of approaching this problem?

Process models as gateway to the algorithmic level

In the field of mathematical psychology, the focus is on the algorithmic level. Theories are instantiated by a set of statistical or mathematical operations that ultimately convert a stimulus input to a behavioral response. The metric of a successful model in this field is whether or not the model can be fit to data, and parameters can be recovered properly. Perhaps most importantly is that the model is falsifiable; the ideal situation is that the model makes narrow predictions about the distribution of behavioral data, and these predictions should be supported by experiments (Roberts and Pashler, 2000). As the focus of this field is at the algorithmic level, the details of the implementational level are abstracted out. That is, the details of how the stimulus maps to a specific mechanism in the model is glossed over for the sake of parsimony. While one could perceive of this tradition as being a shortcoming, abstracting the implementation level allows researchers in the community to focus on generalizing mechanisms, such as across task manipulations (e.g., speed-accuracy trade-off), stimuli (e.g., dots versus bricks), or cognitive processes (e.g., perceptual decision making or memory retrieval).

Despite the powerful advantages cognitive models provide, like DNNs, they are focused on one level of analysis. For example, cognitive models are not well connected to the biological properties of the system they are designed to study. Indeed, while the abstract mechanisms assumed by cognitive models provide parsimony, perhaps making the abstractions more concrete could provide deeper insight into the cognitive process.

* Corresponding author.
E-mail address: buforstmann@gmail.com.

https://doi.org/10.1016/j.neuroimage.2017.12.078
Received 28 September 2017; Accepted 22 December 2017
Available online 30 December 2017

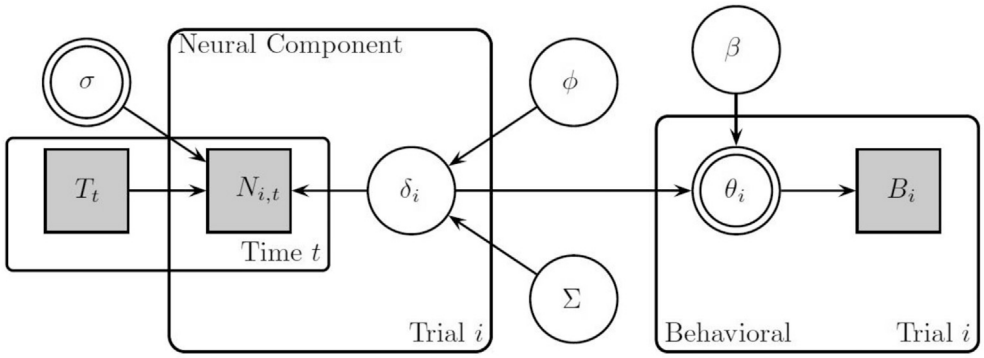

Fig. 1. Graphical diagram of a directed joint model. The left side of the diagram details how neural data *N* change over time, modulated by a set of model parameters. The information extracted from the neural signal is then provided to a computational model, that is then used to make predictions about behavioral data *B*.

Constraining cognitive models with neural data

In an exciting new development, computational modeling has progressed by linking the abstractions assumed by cognitive models directly to patterns of activation in neural data. The goal of this discipline, often referred to as "model-based cognitive neuroscience", is to use trial- or subject-level information about the brain to assess, constrain, or even replace some components of the model (Forstmann et al., 2011; Forstmann and Wagenmakers, 2015; Palmeri et al., 2017). While there are many approaches for linking a stream of neural data to a cognitive model of interest (cf. Turner et al., 2017a,b), one theoretically agnostic approach is to simply treat the neural data as a covariate when fitting the cognitive model to data. As an example, Fig. 1 shows a graphical diagram of a "directed" joint model (Turner et al., 2013, 2015, 2017a,b). Neural data, such as the BOLD response is first modeled on the left side of the diagram, and this information serves as an input term to the cognitive parameter θ, which in turn makes a prediction about the behavioral data. Although the neural model parameters are directly not influenced by the behavioral data B, the neural and behavioral model parameters are implicitly connected by virtue of forming one cohesive model of all observed variables.

When fitting a joint model to data, the estimation of the model parameters is affected by both the neural and behavioral data. In this way, the neural data provide an extra layer of constraint on the behavioral model parameters, and in turn, the behavioral data can provide extra constraint on the neural model parameters. Hence, the joint model shown here bridges both the implementation and algorithmic levels by first articulating the activation or connectivity of the brain data, and then using this information to drive important model parameters.

Using DNNs as input

The approach we advocate is a blend of the two approaches, those that are implementation-oriented and those that are algorithmic-oriented. For example, while the strength of sensory evidence is typically abstracted out in cognitive models, methods like DNN take seriously the computations required to extract information from a visual stimulus. While progress can certainly be made at one level or another, to further our understanding of how the mind is realized within a brain, we must span all levels of analysis. Yet, nothing about DNN or cognitive models prohibits them from being integrated harmoniously. To be sure, DNNs could serve as the impetus for grounding and constraining the abstractions assumed by cognitive models. For example, a DNN applied to a visual stimulus could provide a read out of parameters such as the "drift rate" assumed by models like the Diffusion Decision Model (Ratcliff, 1978). In this way, it might be possible to link DNNs to cognitive models, enabling a more complete understanding of many cognitive processes.

Conclusions

As cognitive computational neuroscientists we live in exciting times. While we can look back on decades of careful and rich cognitive model development, more recent times allow testing these models jointly with sophisticated brain measurements including, e.g., ultra-high resolution magnetic resonance imaging (MRI). The future will show whether such joint modeling approaches can be extended to the implementational level and provide us with a deeper mechanistic understanding of the human mind and brain.

References

Bonini, C.P., 1963. Simulation of Information and Decision Systems in the Firm. Prentice-Hall, Englewood Cliffs, N. J.
Forstmann, B.U., Wagenmakers, E.-J. (Eds.), 2015. An Introduction to Model-based Cognitive Neuroscience. Springer, New York.
Forstmann, B.U., Wagenmakers, E.-J., Eichele, T., Brown, S., Serences, J., 2011. Reciprocal relations between cognitive neuroscience and cognitive models: opposites attract? Trends Cognit. Sci. 6, 272–279.
Marr, D., 1982. Vision: a Computational Investigation into the Human Representation and Processing of Visual Information. Freeman, New York.
Palmeri, T.J., Love, B.C., Turner, B.M., 2017. Model-based cognitive neuroscience. J. Math. Psychol. 76, 56–64.
Ratcliff, R., 1978. A theory of memory retrieval. Psychol. Rev. 85, 59–108.
Roberts, S., Pashler, H., 2000. How persuasive is a good fit? Psychol. Rev. 107, 358–367.
Turner, B.M., Forstmann, B.U., Love, B.C., Palmeri, T.J., Van Maanen, L., 2017a. Approaches to analysis in model-based cognitive neuroscience. J. Math. Psychol. 76, 65–79.
Turner, B.M., Forstmann, B.U., Steyvers, M., 2017b. Simultaneous Modeling of Neural and Behavioral Data. Springer, New York.
Turner, B.M., Forstmann, B.U., Wagenmakers, E.J., Brown, S.D., Sederberg, P.B., Steyvers, M., 2013. A Bayesian framework for simultaneously modeling neural and behavioral data. Neuroimage 72, 193–206.
Turner, B.M., Van Maanen, L., Forstmann, B.U., 2015. Informing cognitive abstractions through neuroimaging: the neural drift diffusion model. Psychol. Rev. 122, 312–336.