Challenges and Approaches to an Information-Theoretic Framework for the Analysis of Embodied Cognitive Systems

Anonymous Author(s)
Affiliation
Address
email

Abstract

Information theory has a long track record as a widely adopted framework to study complex systems. This remains true within the context of cognitive systems, where it has been utilized in psychology, neuroscience, artificial intelligence, artificial life etc. A crucial aspect to what makes a system cognitive is the fact that they are in continuous closed-loop interaction with their environments. This brings certain challenges to utilizing information theory in this context: these systems are multivariate, information has temporal dynamics, information could be distributed across the system and finally, cognition doesn't have to be limited to the brain. In this article, we provide perspectives on these challenges, explain their significance, provide examples of how they have been tackled in other work, and outline the open challenges.

1 Introduction

Although there is a vast and growing amount of data about the neuroanatomy, the neurophysiology, and the behavior of several model organisms, how cognition works is still poorly understood. Much research has been dedicated to understanding how information flows through neural systems. However, with the growing recognition of the central roles that both embodiment and situatedness play in generating and modulating neural activity, the challenge is even more difficult: to understand how behavior is grounded in the dynamics of an entire brain-body-environment system. Such a challenge demands the development of an information-theoretic framework for the analysis of embodied cognitive systems.

By an embodied cognitive system, we mean a dynamic recurrent neural network, coupled to a body, which in turn is coupled to an environment, such that together the coupled brain-body-environment dynamical system produces adaptive behavior. An understanding of an embodied cognitive system entails the characterization of how the interactions between the individual components that make up the neural system, the body, and the environment give rise to the adaptive behavior. In particular, the goal of an informational theoretic analysis must be to characterize how information about the relevant features of the task flow through the complete brain-body-environment system in ways that are relevant for the generation of the adaptive behavior at hand.

There are four main challenges that an informational characterization of a brain-body-environment system must meet. First, it must be able to quantify multivariate interactions. Second, it must be able to quantify how information in the system changes dynamically over time. Third, it must be able to quantify what specific aspects of the features of the task the information is about. Finally, it must be

Submitted to Information-Theoretic Principles in Cognitive Systems Workshop at the 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

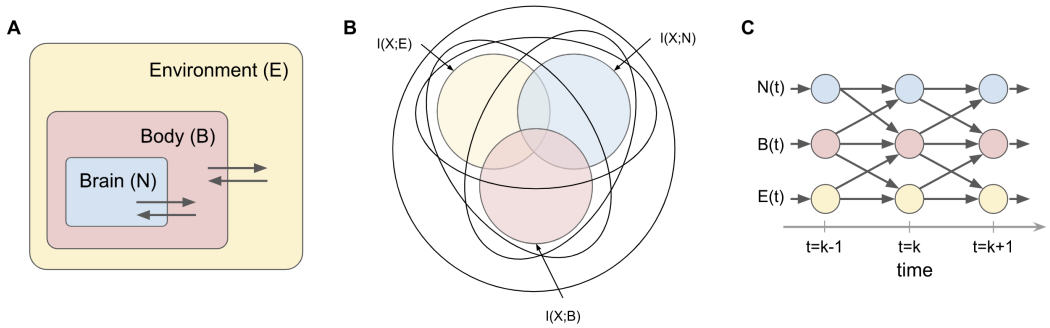

Figure 1: Information theory for Brain-Body-Environment Systems. [A] Illustration of BBE system showing the brain interacting with the body and the body with the environment. Note information flows between each of these layers bidirectionally and thus sets up a closed-loop system. [B] Decomposition of multivariate information in the BBE system about task-relevant variable X. Each inner circle represents information each component (brain, body, environment) has about X. The overlapping regions represent redundant information across them. The non-overlapping region represents unique information and the region outside the inner circles represent the synergistic information. [C] Temporal interaction in a BBE system demonstrating reason for information dynamics. Each component (brain, body, environment) is a dynamical system and when unrolled in time shows the continuous interaction between them. Each node in this graph can be modeled as a random variable to study information (decompostion) at every point in time, and hence its dynamics.

able to quantify information flow in a brain-body-environment system which is coupled in a closed loop.

Of these four challenges for an information-theoretic framework to analyze embodied cognitive systems, the first three have received more attention than the last. This has been in great part driven by the typical open-loop paradigm employed in experimental settings (i.e. feed-forward input-output setup). As such, progress in this area has been driven primarily through the use of synthetic data from simulation models. However, given experimental progress, increasingly model organisms are being studied in a closed-loop fashion (perception drives action and vice versa). In this paper, we discuss each of these challenges and the approaches that have been taken to overcome them to date.

2 Multivariate information

An agent embedded in an environment that is in closed-loop interaction with the environment (perception informs actions which informs perception, and so on), can be modeled as a recurrent dynamical system of recurrent dynamical components i.e. the brain, the body and the environment. To understand the role of one of those components, one common approach is to perform an ablation study of that unit. While ablation experiments provide conclusive evidence of that units participation/requirement in the behavior (see note in appendix about information ablation), they do not have any explanatory power. To this end Partial Information Decomposition (PID) [Williams and Beer, 2010a] is an information theoretic framework that allows decomposing the total information across multiple sources (think brain, body, or environment) about a target variable (think task-relevant variable) into unique information that each source about the target, redundant information that may exist across multiple sources, as well as synergistic information that is only available from referring to multiple sources at the same time and not from just one source alone. Consider two sources of information (two random variables) $X1$ and $X2$, about the random variable Y. If $X = \{X1, X2\}$, the total mutual information $I(X;Y)$ can be written as follows

$$I(X;Y) = U(X1;Y) + U(X2;Y) + R(X;Y) + S(X;Y) \tag{1}$$

where U denotes unique information, R denotes redundant information and S denotes synergistic information from these sources about Y. Naturally, with more than two sources, redundant and synergistic information will be available for the all combinations of the different sources. The decomposition can be better understood when visualized as a Venn diagram as shown in figure 1B

for three sources: brain, body and environment. Also, note that each random variable used in these descriptions is multi-dimensional.

This approach could be applied simply within each component of the BBE system. For instance, within just the brain, all neurons feeding into another neuron can be analyzed a set of sources feeding information into the target neuron about the variable of interest. This allows us to evaluate the unique, redundant and synergistic information that the target neuron receives from the source neurons about the variable of interest. Furthermore, when these variables are offset in time, as shown in figure 1C, partial information decomposition has been shown to equivalent to transfer entropy [Schreiber, 2000, Williams and Beer, 2011]. In neuroscience, PID has been utilized to study task relevant changes in fMRI [Lizier et al., 2011, Anzellotti and Coutanche, 2018], information distribution dissociated cultures [Sherrill et al., 2021] and importantly in neural models of BBE systems [Beer and Williams, 2015].

While PID scales well for high-dimensional systems, open challenges with analyzing multivariate systems involves scaling up the number of sources of information. If one were to decompose the BBE system into a three source PID framework as shown in figure 1B, each source can be any dimensional random variable. However, if one were to decompose it further into the individual sub-units within each of those sources, then PID would not scale well (imagine what a 7 source Venn diagram would look like).

3 Dynamic information

The source of information in dynamical systems can quickly get obscured in time [Amblard and Michel, 2012]. In a system where the agent and environment are continuously interacting, a signal that arises in the environment (or agent) quickly gets reflected in the agent's (or environment's) dynamics thus making it impossible to disentangle the source. Candadai and Izquierdo [2020] show this phenomenon in a simulated setting where the environment drives a dynamical neural network model or vice versa (i.e. even when the behavior does not involve mutual interaction between agent and environment). Consequently, information quantities measured across time become estimates of the generalized correlation between those two random variables and does not reveal the direction of information flow i.e. causation. Thus, it becomes crucial to study temporal fluctuation in information across components of the BBE system to track the source of information. In other words, unrolling information in time allows us to identify *when* specific milestones occur over the course of a task and *where* they are initiated.

Information theoretic measures are readily adaptable to unroll them over time. This involves modeling variables of interest as random processes rather than random variables. In other words, a variable of interest at specific temporal landmarks of a task (or even better, along each time point of the task) are modeled as random variables and information theoretic analyses are performed at those time points by estimating probability densities at these time points (each node the temporal graph shown in figure 1C would be a random variable). This has been applied in a few different contexts: Williams and Beer [2010b] demonstrate how information measured over time allows us to identify when a decision about catch/avoid is made in an artificial agent optimized to perform a relational categorization task; Izquierdo et al. [2015] show the dynamics of information in a computation model of C. *elegans*, demonstrating how information about the chemosensory stimulus flows through its neural circuitry over time. Based on the graph in figure 1C, information at time t in the brain about task-relevant variable X would include two other sources of information $N(t-1)$ and $B(t-1)$. Using the PID framework discussed in the previous section, the decomposition of multivariate information of these three sources allows us to study the origin and flow of information across the BBE system components in time.

Open challenges in temporal analysis of information quantities rests on the experimental side; accurate measurements of variables of interest at specific landmarks during tasks over multiple trials (to build a distribution) is challenging. However, with increasingly advanced technology becoming available, such experiments are starting to become viable thus further increasing the scope for application of information theory [Yawo et al., 2013, Wang et al., 2011].

4 Specific information

Information about a task-relevant variable is often expected to be stored at a specific location within a BBE system. However, it has been shown that it could in fact be available in a decentralized fashion. For instance, population coding could be regarded as distributed where some neurons store information about a certain slice of the overall distribution of the variable of interest whereas other neurons could encode information about other slices. In other words, measuring information about *specific* values of the variable of interest would enable acquiring a more granular explanation of the operation of a BBE system (X in figure 1B, could refer to only a specific value of that task relevant variable). This involves, as one might expect, building a distribution of the BBE variables in response to several trials of a particular value (or a set of values) of the task-relevant variable, and then simply measuring the same information quantities such as mutual information. Ma et al. [2017] and Steinmetz et al. [2019] are examples of studies where spatial distribution of neural encoding was studied but all data required to perform a specific information study was available. One challenge in this regard, is to establish this methodology more concretely so that these opportunities are taken advantage of.

5 Embodied information

An agent embedded in an environment that is in closed-loop interaction with the environment (perception informs actions which informs perception, and so on), can be modeled as a recurrent dynamical system of recurrent dynamical components i.e. the brain, the body and the environment. To understand the role of one of those components, namely the brain, in such a setting requires answering questions along the lines of: Where in the system does information about X originate? When does it originate? Where does it exist at a given point in time? Answering these questions brings together the three previous sections outlined above in BBE systems. Importantly, when dealing with embedded, embodied systems the brain does not receive special treatment. It is part of a larger dynamical system. While, the brain may be the object of interest, every variable in the BBE system is a potential candidate for storing information. A great example of this effect is the relational categorization task studied in Beer and Williams [2015]. The task's objective is for as simulated agent to "catch" an object if the previous object it was provided was larger than the current object and "avoid" it if wasn't. Information theoretic analyses of the optimized agents as a BBE system revealed that information about the size of the previous object wasn't stored in the brain (neural network activations). The agent offloaded that information to the environment by moving itself to different positions in space depending on the size of the previous object. This was revealed by measuring the information about the first object's size in the neural network and comparing it to information in the spatial location of the agent after it had perceived the first object. Essentially,

6 Conclusion

In summary, this paper outlines the four major aspects of information theory that aid in furthering our understanding of adaptive behavior. The discussion here and references below goes to show that in essence, information theory already provides the generic tools necessary for building a much better understand of adaptive behavior. It is the application of those tools and the scope within which they are applied that require more careful thought and consideration to take advantage of the full potential of these tools. Having said that, there is continuing research that help expand the field of information for better estimators, better ways of studying polyadic interactions, better ways of scaling them computationally and so on. The important message is that the study of cognition does not necessarily have to wait for those developments and could certainly benefit from adopting these tools more widely today (we provide references to some software tools in the appendix).

References

Paul L Williams and Randall D Beer. Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:1004.2515*, 2010a.

Thomas Schreiber. Measuring information transfer. *Physical review letters*, 85(2):461, 2000.

Paul L Williams and Randall D Beer. Generalized measures of information transfer. *arXiv preprint arXiv:1102.1507*, 2011.

Joseph T Lizier, Jakob Heinzle, Annette Horstmann, John-Dylan Haynes, and Mikhail Prokopenko. Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fmri connectivity. *Journal of computational neuroscience*, 30(1):85–107, 2011.

Stefano Anzellotti and Marc N Coutanche. Beyond functional connectivity: investigating networks of multivariate representations. *Trends in cognitive sciences*, 22(3):258–269, 2018.

Samantha P Sherrill, Nicholas M Timme, John M Beggs, and Ehren L Newman. Partial information decomposition reveals that synergistic neural integration is greater downstream of recurrent information flow in organotypic cortical cultures. *PLoS computational biology*, 17(7):e1009196, 2021.

Randall D Beer and Paul L Williams. Information processing and dynamics in minimally cognitive agents. *Cognitive science*, 39(1):1–38, 2015.

Pierre-Olivier Amblard and Olivier JJ Michel. The relation between granger causality and directed information theory: A review. *Entropy*, 15(1):113–143, 2012.

Madhavun Candadai and Eduardo J Izquierdo. Sources of predictive information in dynamical neural networks. *Scientific reports*, 10(1):1–12, 2020.

Paul L Williams and Randall D Beer. Information dynamics of evolved agents. In *International Conference on Simulation of Adaptive Behavior*, pages 38–49. Springer, 2010b.

Eduardo J Izquierdo, Paul L Williams, and Randall D Beer. Information flow through a model of the c. elegans klinotaxis circuit. *PloS one*, 10(10):e0140397, 2015.

Hiromu Yawo, Toshifumi Asano, Seiichiro Sakai, and Toru Ishizuka. Optogenetic manipulation of neural and non-neural functions. *Development, growth & differentiation*, 55(4):474–490, 2013.

Kaiyu Wang, Yafeng Liu, Yiding Li, Yanmeng Guo, Peipei Song, Xiaohui Zhang, Shaoqun Zeng, and Zuoren Wang. Precise spatiotemporal control of optogenetic activation using an acousto-optic device. *PLoS One*, 6(12):e28468, 2011.

Chaolin Ma, Xuan Ma, Jing Fan, and Jiping He. Neurons in primary motor cortex encode hand orientation in a reach-to-grasp task. *Neuroscience Bulletin*, 33(4):383–395, 2017.

Nicholas A Steinmetz, Peter Zatka-Haas, Matteo Carandini, and Kenneth D Harris. Distributed coding of choice, action and engagement across the mouse brain. *Nature*, 576(7786):266–273, 2019.

Madhavun Candadai and Eduardo J Izquierdo. infotheory: A c++/python package for multivariate information theoretic analysis. *arXiv preprint arXiv:1907.02339*, 2019.

Ryan G James, Christopher J Ellison, and James P Crutchfield. dit: a python package for discrete information theory. *Journal of Open Source Software*, 3(25):738, 2018. doi: 10.21105/joss.00738.

Patricia Wollstadt, Joseph T. Lizier, Raul Vicente, Conor Finn, Mario Martınez-Zarzuela, Pedro Mediano, Leonardo Novelli, and Michael Wibral. Idtxl: The information dynamics toolkit xl: a python package for the efficient analysis of multivariate information dynamics in networks. *Journal of Open Source Software*, 4(34):1081, 2019. doi: 10.21105/joss.01081.

Joseph T Lizier. Jidt: An information-theoretic toolkit for studying the dynamics of complex systems. *Frontiers in Robotics and AI*, 1:11, 2014. doi: 10.3389/frobt.2014.00011.

Michael Lindner, Raul Vicente, Viola Priesemann, and Michael Wibral. Trentool: A matlab open source toolbox to analyse information flow in time series data with transfer entropy. *BMC neuroscience*, 12(1):1–22, 2011.

A Appendix

A.1 Information ablation

Information in natural systems is often encoded in the distribution of a signal rather that just its presence. To this end, to better mimic absence of information from a particular source, rather then completely remove any signal coming from that source we could manipulate the source signal distribution. A natural way to achieve this between two neurons would be to replace the signal from source neuron to target neuron to be equal to the source neuron's activity's expected value rather than 0. This helps remove confounds that physical lesions can bring, such as in situations where a source neuron while providing no information could still act as a constant clamp-like input to a target neuron to drive the behavior - physical lesions alone would not be able to detect such behavior.

A.2 Software packages for information theoretic analyses

1. infotheory - C++/Python; PID measures as well as transfer entropy estimation for upto 3 sources [Candadai and Izquierdo, 2019]
2. dit - Python; PID for discrete variables [James et al., 2018]
3. IDTxL - Python; PID and transfer entropy for upto 2 sources [Wollstadt et al., 2019]
4. JDIT - JAVA; primarily designed for measuring transfer entropy but also includes other measures such as entropies and mutual information [Lizier, 2014]
5. TRENTOOL - MATLAB toolbox which primarily caters to analog neural data such as MEG [Lindner et al., 2011]