Neural Networks 103 (2018) 94–107

Contents lists available at ScienceDirect

Neural Networks

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

Information-theoretic decomposition of embodied and situated systems

Federico Da Rold

School of Computing, Electronics and Mathematics, Plymouth University, Plymouth PL4 8AA, UK

ARTICLE INFO

Article history:
Received 13 July 2017
Received in revised form 1 January 2018
Accepted 14 March 2018
Available online 27 March 2018

Keywords:
Embodied cognition
Situated cognition
Evolutionary robotics
Information theory
Antireductionism

ABSTRACT

The embodied and situated view of cognition stresses the importance of real-time and nonlinear bodily interaction with the environment for developing concepts and structuring knowledge. In this article, populations of robots controlled by an artificial neural network learn a wall-following task through artificial evolution. At the end of the evolutionary process, time series are recorded from perceptual and motor neurons of selected robots. Information-theoretic measures are estimated on pairings of variables to unveil nonlinear interactions that structure the agent–environment system. Specifically, the mutual information is utilized to quantify the degree of dependence and the transfer entropy to detect the direction of the information flow. Furthermore, the system is analyzed with the local form of such measures, thus capturing the underlying dynamics of information. Results show that different measures are interdependent and complementary in uncovering aspects of the robots' interaction with the environment, as well as characteristics of the functional neural structure. Therefore, the set of information-theoretic measures provides a decomposition of the system, capturing the intricacy of nonlinear relationships that characterize robots' behavior and neural dynamics.

1. Introduction

Traditional cognitive science conceives the mind as an amodal computational machine whose function is manipulating abstract symbols according to a set of logical rules (Fodor, 1975; Newell & Simon, 1976; Pylyshyn, 1986). Therefore, intelligence is essentially high level and discrete reasoning, whereas perception and action, body and environment, are auxiliary appendices unnecessary for understanding cognitive processing. However, such algorithmic view of intelligence is challenged by the embodied cognition perspective (see for example Pfeifer & Bongard, 2006, Varela, Thompson, & Rosch, 1997), which poses as fundamental pillar of human thinking the constant interaction of the biological cognizant with the surrounding physical environment throughout a continuous flow of sensorimotor loops (Beer, 2008; Thelen & Smith, 1996). Accepting such an integrative idea of cognition leads, however, to significant problems in the analysis of embodied and situated systems. In fact, current approaches are mainly based on a qualitative geometrical description of the systems' behavior (see for example Beer, 1995). Although visual descriptions have the merit of explaining the evolution of agent–environment interaction thorough time, a precise numerical explanation of observed phenomena is an essential prerequisite to advancing a scientific field.

To overcome this issue, we attempt to analyze an evolutionary robotic model (Nolfi & Floreano, 2000) performing a wall-following task with the tools developed in the field of information theory. The choice of such cognitive model is motivated by its consistency with an operationalization of the embodied and situated view. In the experimental scenario, a robot, capable of receiving perceptual inputs and producing motor responses, is controlled by an artificial brain, and the agent autonomously develops a solution to a problem defined by the surrounding environment. In order to avoid a disentanglement of the sub-components of the system, this work is framed within the context of an antireductionist philosophy of science (Ahn, Tewari, Poon, & Phillips, 2006a, b; Fang & Casadevall, 2011) and system science (Von Bertalanffy, 1968), using concepts recently developed in the field of system biology (Basso et al., 2005; Margolin, Wang et al., 2006). Information-theoretic measures do not directly accommodate into an antireductionist scientific framework, as for example chaos theory (Da Rold, 2017). However, we apply such measures to all possible pairings of connected variables, thus unveiling the intricate web of dynamical and structural relationships of the indivisible system brain–body–environment. Following these principles, we propose the idea of an information-theoretic decomposition, which refers to a mapping of the nonlinear relationships among sub-components of the system for explaining the robot's behavior from an integrative perspective.

E-mail address: federico.darold@plymouth.ac.uk.

https://doi.org/10.1016/j.neunet.2018.03.011

1.1. Related works

Information theory has already been applied to robotic platforms to unveil how embodied and situated agents structure the information, showing that different behaviors (Tarapore, Lungarella, & Gómez, 2004), objects (Gomez, Lungarella, & Tarapore, 2005), body morphology and learning stages (Lungarella & Sporns, 2006) affect the outcome of information-theoretic measures. The transfer entropy, a measure that estimates the direction of the information flow between pairs of variables, detects different relationships among perceptual and motor variables if a quadruped robot exhibits different behaviors or acts in arenas with different grounds that vary the friction (Schmidt, Hoffmann, Nakajima, & Pfeifer, 2013). The explanatory power of information theory for quantitatively explaining behavioral patterns autonomously developed in robotic systems is further investigated by applying predictive information and excess entropy, thus measuring the information contained in past states about the future (Martius & Olbrich, 2015). Yamada, Nishikawa, Shida, Niiyama, and Kuniyoshi (2011) propose an information-theoretic analysis of a quadruped robot with biological plausible muscles, tendons, and neural controller. The information flow and structure estimated on data collected from perceptual and motor variables show a significant difference in robots equipped either with mono or bi-articular muscles. The application of mutual information to embodied and situated systems is extended to a human–robot interaction scenario (Stoelen, de Tejada, Huete, Balaguer, & Bonsignorio, 2015) for developing a set of measures aimed at characterizing real-world applications of adaptive systems. The information exchange between a neuro-robotic platform and human participants during the interaction is quantified with the mutual information, providing a sound and reliable benchmark. An animal robot interaction scenario proves that the transfer entropy correctly indicates that a zebrafish is more influenced by a replica than a conspecific (Butail, Ladu, Spinello, & Porfiri, 2014).

Information-theoretic approaches have been applied to biological neural networks to unveil functional relationships among neuronal ensembles (Friston, Moran, & Seth, 2013). In fact, transfer entropy is a consistent operationalization of the definition of causation proposed by Wiener (1956). Following this perspective, Wibral, Vicente, and Lizier (2014) show that this information-theoretic measure identifies true interactions among time series recorded from human participants involved in a simple motor task using magnetoencephalography. Further studies on a perceptual closure task demonstrate that the transfer entropy is capable of reconstructing information patterns between brain areas that are confirmed by neurophysiological studies (Wollstadt, Martínez-Zarzuela, Vicente, Díaz-Pernas, & Wibral, 2014). Furthermore, Buehlmann and Deco (2010) study phase synchronization in a simulated brain model consisting of integrate-and-fire neurons. Results show that the information transfer is correlated to the phase relation, thus providing evidence to the communication through coherence hypothesis (Fries, 2005).

1.2. Proposed approach

An approach closely related to the work described in this article is applied to a model that simulates a minimal klinotaxis neural circuit of the nematode *c. elegans* (Izquierdo, Williams, & Beer, 2015). By calculating mutual information and transfer entropy, the authors describe the complete loop of information flow from environmental stimuli, to sensory and hidden neurons, to motor neurons and actuators. In a related work, Beer and Williams (2015) analyze a minimal cognitive agent involved in a relational categorization task, outlining a proposal for relating dynamical systems and information theory, showing the complementarity and consistency of the two styles of explanations.

In this article, we study a minimal neuro-robotic system performing a wall-following task, thus creating a cognitive model characterized by a restricted and controllable number of variables. However, such models are widely employed for testing hypothesis concerning cognitive functions (De La Cruz, Di Nuovo, Di Nuovo, & Cangelosi, 2014) and novel analytical methodologies (Beer & Williams, 2015).

The proposed information-theoretic framework introduces the following innovations. Firstly, the embodied system is analyzed estimating local information-theoretic measures, which evaluate the information of specific measurements during the execution of the task. Therefore, the outcome of the local form consists of a time series describing the dynamic of information flow and exchange, rather than providing an average or expected value of the entire dataset. In this respect, our work suggests an alternative for bridging an analysis based on dynamical system theory and information-theoretic approaches (Beer & Williams, 2015). Secondly, major variables such as body morphologies or the environment are not modified. The only variability between repetitions of the experiment are different values of the synaptic weights of the neural controller, leading to minor behavioral differences. Therefore, the focus of the proposed analysis is to assess the capability of information-theoretic measures to detect subtle changes in the embodied and situated system. Thirdly, a neuro-robotic system provides a concrete instantiation of communication channel, i.e. the artificial synapses. By lesioning connections between neurons, the inferences and hypotheses that follow the information-theoretic analysis are confirmed or rejected based on empirical verification.

2. Materials and methods

In this section we report the details of the experimental scenario and the measures utilized during data analysis. Evolutionary robotics is a semi-supervised learning method that autonomously explores the solution space defined by a global utility function. By mimicking the evolutionary process, populations of robots develop a solution finding the optimal free parameters of the system, adjusting the values of synaptic weights and motor biases. At the end of evolution, data is recorded from the 4 neurons of the neural network during the execution of the task. The recorded time series are subsequently analyzed with information-theoretic measures.

2.1. Simulated robotic scenario

Populations of simulated e-puck robots are evolved using Evorobot, a scientific software developed for evolutionary robotics experiments (Nolfi & Gigliotta, 2010). The simulated agents are miniature cylindrical robots of 7 cm diameter with two differential wheels and 8 infrared sensors placed around the body (Mondada et al., 2009). In this experiment, only the front middle left and front middle right sensors are activated and connected to the neural controller (Fig. 1(b)).

The environment is a simple squared maze with inner walls of 30 cm length and outer walls of 54 cm, shaping 4 corners of 90°. At the beginning of the trial robots are placed after the bottom-right corner, as shown in Fig. 1(a), with the front part of the agent directed toward the upper-right corner. The aim of the simulation is to mapping the relations between variables forming the system, rather than the robots' ability at learning multiple tasks or developing robustness toward noise. Therefore, the environment where robots evolve is not manipulated to avoid a combinatorial explosion of the factors involved during the information-theoretic analysis. In fact, the length and displacement of the walls, as well as the starting position, are constant. During the learning stage, the lifespan of a robot is limited to 2000 time steps and a collision