

Predicate learning via neural oscillations supports one-shot generalization between video games

Leonidas A. A. Doumas (alex.doumas@ed.ac.uk)

Department of Psychology University of Edinburgh, 7 George Square
Edinburgh, EH8 9JZ United Kingdom

Guillermo Puebla (guillermo.puebla@ed.ac.uk)

Department of Psychology University of Edinburgh, 7 George Square
Edinburgh, EH8 9JZ United Kingdom

John E. Hummel (jehummel@illinois.edu)

Department of Psychology University of Illinois, 901 West Illinois Street
Urbana, Illinois 61801, USA

Andrea E. Martin (andrea.martin@mpi.nl)

Max Planck Institute for Psycholinguistics, Wundtlaan 1
Nijmegen, 6525XD The Netherlands

Abstract

Humans readily generalize, applying prior knowledge to novel situations and stimuli. Advances in machine learning have begun to approximate and even surpass human performance, but these systems struggle to generalize what they have learned to untrained situations. We present a model based on well-established neurocomputational principles that demonstrates human-level generalization. This model is trained to play one video game (Breakout) and performs one-shot generalization to a new game (Pong) with different characteristics. The model generalizes because it learns structured representations that are functionally symbolic (viz., a role-filler binding calculus) from unstructured training data. It does so without feedback, and without requiring that structured representations are specified a priori. Specifically, the model uses neural co-activation to discover which characteristics of the input are invariant and to learn relational predicates, and oscillatory regularities in network firing to bind predicates to arguments. To our knowledge, this is the first demonstration of human-like generalization in a machine system that does not assume structured representations to begin with.

Keywords: predicate learning; generalization; neural networks; symbolic-connectionism; neural oscillations

Introduction

Recently deep neural network (DNN) systems have reached and even exceeded human levels of performance on a range of cognitive tasks (for a review see, Hassabis, Kumaran, Summerfield, & Botvinick, 2017). For example, DNNs have learned to master an impressive number of games (Mnih et al., 2015; Silver et al., 2017). DNNs are general, in that they can learn to perform a variety of tasks without a priori background knowledge. Nevertheless, while DNNs readily perform interpolation (i.e., generalization to untrained items from within the bounds

of the training set), they struggle to perform extrapolation (i.e., generalization to items from outside the bounds of the training set). For example, a network trained to play Breakout must be completely retrained to play Pong (Mnih et al., 2015).

In contrast, a person is able to quickly catch on to playing a game like Pong after learning to play a game like Breakout. After all, Breakout and Pong are very similar: In both games the objective is to use a paddle to keep a ball in play, and to hit the ball toward some goal. While in Breakout the ball is played vertically towards blocks at the top of the screen, and in Pong the ball is played horizontally towards an opponent paddle.

Accounts of how humans generalize are frequently based on powerful symbolic languages that include structured relations (or predicates), which can be promiscuously applied to new arguments (Anderson, 2009; Doumas & Hummel, 2012; Lake, Ullman, Tenenbaum, & Gershman, 2017). In this view, we have abstract representations like right-of and above. These representations allow us to characterize different domains with the same representations, and generalize what we have learned about these representations across domains. Structured models, however, face a challenge that is complementary to that which DNNs face: They characteristically require the modeler to specify a collection of necessary representational structures in advance of any actual learning (e.g. Lake, Salakhutdinov, & Tenenbaum, 2015).

We have previously proposed a neural network model of how structured representations are instantiated in a biologically plausible neural system, and how such representations are learned in the first place (Doumas, Hummel, & Sandhofer, 2008). The model, called DORA, uses unsupervised comparison to discover which characteristics of the input are invariant, and to learn functional predicates; it then applies these predicates to arguments in a symbolic fashion, using oscillatory regularities to dynamically bind predicates and arguments. DORA learns representations that are functionally and formally symbolic from flat vector data, without feedback, and



without requiring that structured representations be specified a priori.

In the following we show that after learning to play one video game, Breakout, the representations that DORA learns support generalisation to a completely new game, Pong, in one shot. Importantly, DORAs learning and reasoning rely intimately on the phase dynamics that carry binding information in the model.

Model description

DORA is a symbolic-connectionist model descended from LISA (Hummel & Holyoak, 2003). Its operation is summarized as follows. (1) DORA starts with representations of differentiated objects encoded as flat feature vectors. (2) Through a process of analogical mapping, objects are compared (and co-activated) and their feature vectors are superimposed. (3) DORA learns a representation of the overlaid pattern through Hebbian learning. The resulting representation is an encoding of what the compared objects have in common. (4) The learned representations are bound to objects by systematic asynchrony of firing, resulting in functional single-place predicates. (5) Co-occurring sets of single-place predicates are linked to form functional multi-place relations. Below we provide a conceptual overview of DORA's operation at a high level.

Computational macrostructure

DORA has a long-term-memory (LTM; see Fig. 1) composed of bidirectionally connected layers of units. Units in LTM are referred to as token units. Token units in the lowest layer of LTM are bidirectionally connected to a common pool of feature units. Token units are yoked to inhibitors that integrate input from their yoked unit and token units in higher layers, and fire after reaching a threshold. Yoked inhibitors serve the purpose of implementing phasic firing and refractory periods in the token units, which are important for implementing dynamic binding in the network.

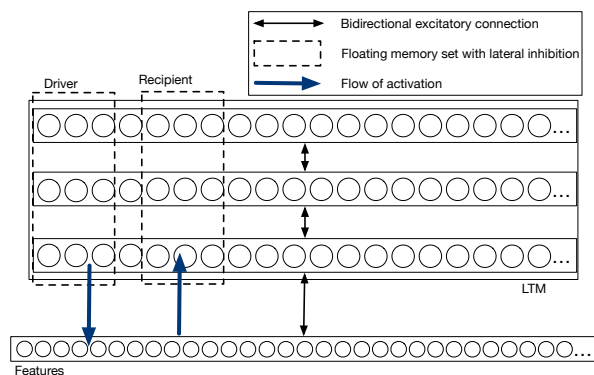


Figure 1: Macrostructure of the DORA network.

Potentiated sets of token units, or memory sets (dashed boxes in Fig. 1), correspond to DORA's working memory.

Memory sets include, the driver, DORAs current focus of attention, and the recipient, DORAs current active memory. Token units in the same layer inhibit one another within, but not across, memory sets. Activation in the model flows from the token units in the driver to token units in the recipient and LTM via the shared pool of feature units.

Neurosymbolic representations

DORA begins with representations of objects coded as flat feature vectors. For example, DORA might represent a ball with a token unit connected to a set of features (see Fig. 2a). In terms of cortical computation, feature nodes can be thought of as aggregate units, perceptual representations, or activation states over networks.

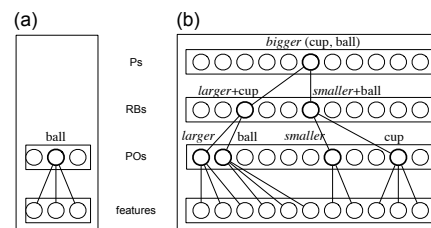


Figure 2: Representations in DORA.

DORA eventually learns representations of a form we call LISAese (Hummel & Holyoak, 2003). Full propositions in LISAese are encoded across layers of units (Fig. 2b). For ease of exposition we label the layers. At the bottom of the hierarchy, are feature units. The layer of token units connected to the features are POs (for predicate-object), followed by RBs (for role-binding) and Ps (for proposition) units. Feature units code for the properties of represented instances in a distributed manner. POs, conjunctively link collections of feature units encoding objects (and, after learning, predicates). RBs conjunctively link POs into role-filler pairs. Ps conjunctively link RBs to form multi-place relational structures. As an example, a LISAese representation of the relational proposition bigger (ball, cup) is depicted in Fig. 2b.

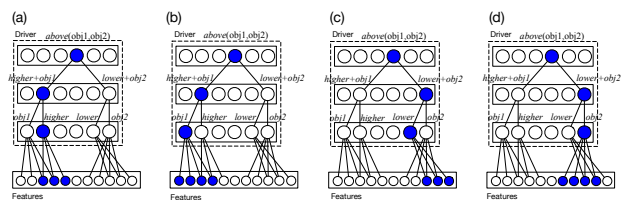


Figure 3: Binding in DORA.

To process a proposition, role-filler bindings must be represented dynamically on the units that maintain role-filler independence (i.e., POs and feature units; see Dumas et al., 2008). In DORA, roles are dynamically bound to their fillers by systematic asynchrony of firing. When laterally inhibitive units

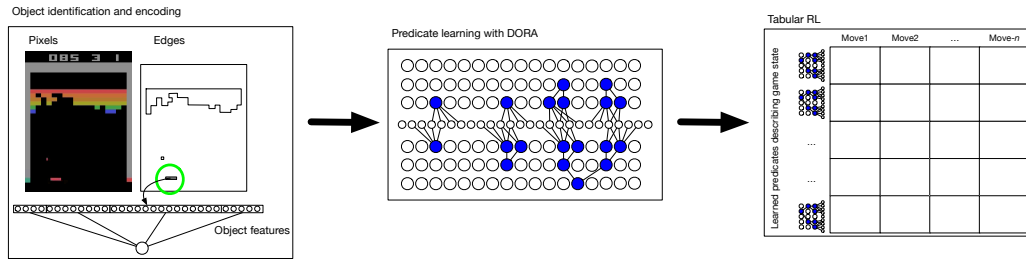


Figure 4: Predicate learning in context.

are linked by a conjunctive node, they will naturally oscillate out of synchrony and in direct sequence when their conjunctive unit becomes active. This emergent oscillatory pattern is an effective binding signal. For example, as a proposition in the driver becomes active, bound roles and objects fire in direct sequence. (see Fig. 3a-d).

Processing

DORA is a settling network. It starts in some state, such as a set of units in driver (e.g., chosen at random from LTM, or based on DORAs current perceptual state such as a video-game screen shot). Token units in the driver compete (via lateral inhibition) to become active, and activation flows to token units in the recipient and LTM via shared feature units. DORA eventually settles into some state (e.g., with some units active in driver and recipient). Due to the refraction of nodes and yoked inhibitors, this state will eventually become upset and the process will start again.

Starting with some units in the driver, DORA cycles through five operations: retrieving, mapping, predicate learning, refining, and generalising. During retrieval, active token units in the driver pass activation to tokens in LTM via shared features. After all tokens in the driver have fired, DORA retrieves (i.e., potentiates) units from LTM into the recipient using the Luce choice rule.

During mapping, DORA discovers structural correspondences between token units in the driver and recipient. As token units in the driver become active, token units in the recipient compete (via lateral inhibition) to become active. DORA learns mapping connections (via a modified Hebbian learning rule) between simultaneously active units in the same layer across driver and recipient.

During predicate learning, DORA learns functional predicate representations of, first, single-place, and, subsequently, multi-place predicate. As token units in the driver become active, they activate corresponding token units in the recipient (through shared features and mapping connections). In response, DORA recruits an unconnected token unit in response to active units in the layer directly below. Connections between units are updated by Hebbian learning. As a result, recruited POs learn connections to the featural overlap of mapped objects, and recruited RBs and Ps learn conjunctive encodings of co-occurring lower-level tokens. The result-

ing collections of units function like single-place predicates, and eventually multi-place relations (as in Fig. 3).

During refinement, DORA learns schematised representations of mapped items in driver and recipient. As token units in the driver become active, they activate corresponding units in the recipient (through shared features and mapping connections). Token units in LTM are recruited to match active mapped units in the driver. Just as with predicate learning, connection weights between units are updated by Hebbian learning. The result is a refined representation in LTM of the mapped representations in driver and recipient.

During generalisation, DORA performs relational generalisation. When unmapped token units in the driver become active, token units are recruited in the recipient. Connection weights between units are updated by Hebbian learning. The result is that unmapped token units (and thus structure) from the driver are essentially copied into the recipient.

Model in context

DORA is a model of representation learning. It assumes that objects are differentiated and makes no strong claims about how choices between available options (i.e., moves in a video game) are made. As such, we situated DORAs predicate learning algorithm between a visual pre-processor, and tabular Q-learning (Watkins, 1989) (see Fig. 4). The visual pre-processor served to differentiate objects, and the tabular Q-learning allowed DORA to learn associations between representational states and move options in a game.

We tested two versions of the visual pre-processor. The first was a pre-trained mask R-CNN (He, Gkioxari, Dollár, & Girshick, 2017), delivering object outlines and delimiting rectangles. The second performed the exact same task but used edge detection (via local contrast) with an inbuilt bias such that any enclosed edges were treated as a single object. As both networks behave identically for the present purposes we used the later because is computationally simpler and faster.

Simulations

We compared (1) an implementation of DORA with Q-learning against (2) DQN; (3) DQN with the same pre-processed inputs used by DORA; (4) a supervised deep neural network (DNN) with the same pre-processed inputs used by DORA with fixed frame skipping; (5) a supervised DNN with the same pre-

processed inputs used by DORA with random frame skipping; (6) Humans (two Breakout and Pong novices). We trained all these systems to play one videogame (Breakout), and then tested their ability to generalize to a different videogame (Pong) without any explicit training. Finally, we evaluated these systems' ability to switch back to playing the original game, after time spent learning to play the second.

For the first 250 games of Breakout, DORA made random moves, generating game states from which it learned structured representations in an unsupervised manner as described above. DORA successfully learned predicate representations encoding to instances such as *more-y*(object1, object2) and *more-x*(object1, object2). DORA then attempted to learn to play Breakout using the representations that it had learned during the first 250 games to represent the current game screen and then made a response. Associations between these learned representations and successful moves were learned via tabular Q-learning.

Fig. 5a shows the performance of all networks on Breakout as an average score of the last 100 games played, and a high score. All systems performed quite well, reaching levels of performance that matched or exceeded human participants. As would be expected, DORA took far fewer games to learn to play Breakout than any of the other networks (1,000 vs. 10,000,000 games for DORA and DQN, respectively).

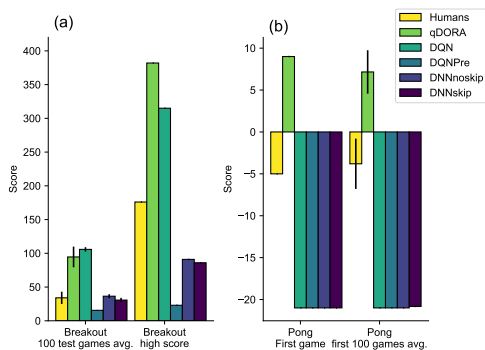


Figure 5: Predicate learning in context.

We then tested the capacity of the networks to play a new videogame, Pong. DORA had learned to play Breakout by learning associations between relational configurations and actions. During its first game of Pong, DORA represented the game state using the relations it had learned playing Breakout. DORA discovered a correspondence between the action sets in the two games: particularly, *more-y/less-y* of the paddle (the paddle moves up and down) in Pong and *more-x/less-x* of the paddle (the paddle moves horizontally) in Breakout. This correspondence allowed DORA to infer via relational generalization the relational configurations that reward specific moves in Pong. For example, just as *more-x*(ball, paddle) tends to reward a *more-x* move of the paddle in Breakout, *more-y*(ball, paddle) rewards a *more-y* move of the paddle in Pong. Fig. 5b shows the performance of the human players and the net-

works on the first game of Pong after training on Breakout and the average performance over the first 100 games playing Pong. Like a human player, DORA performed at a high level on Pong on a single exposure to the game and continued to play Pong at a high level. By contrast, all other networks showed poor performance—which is unsurprising given previous results using DNNs and transferring to different contexts.

General Discussion

We have shown that a machine system can perform extrapolatory generalization through predicate learning. Specifically, DORA used predicate learning to discover symbolic representations from video game screen shots without feedback, and without assuming any structured representations a priori. Crucially, the predicate representations that DORA learned allowed it to extrapolate its knowledge to untrained situations. To our knowledge, this is the first demonstration of human-like generalization, or extrapolation, in a machine system that does not assume structured representations to begin with. Importantly, the solution makes use of well-established neuro-computational principles.

References

- Anderson, J. R. (2009). *How can the human mind occur in the physical universe?* Oxford University Press.
- Doumas, L. A., & Hummel, J. E. (2012). Computational models of higher cognition. *Oxford handbook of thinking and reasoning*, 52–66.
- Doumas, L. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological review*, 115(1), 1.
- Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, 95(2), 245–258.
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961–2969).
- Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological review*, 110(2), 220.
- Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, 350(6266), 1332–1338.
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and brain sciences*, 40.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... others (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529.
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... others (2017). Mastering the game of Go without human knowledge. *Nature*, 550(7676), 354.
- Watkins, C. J. C. H. (1989). *Learning from delayed rewards*. Unpublished doctoral dissertation, University of Cambridge.