Motivation Driven Learning of Action Affordances

Ignasi Cos-Aguilera[⋆†], Lola Cañamero[†] and Gillian M. Hayes[⋆]

[⋆]IPAB, School of Informatics, University of Edinburgh,
Mayfield Road, JCMB 1.1106, EH9 1JZ Edinburgh, Scotland, UK.
ignasi@dai.ed.ac.uk, gmh@inf.ed.ac.uk

[†]School of Computer Science, University of Hertfordshire,
College Lane, Hatfield, Herts, AL10 9AB, UK.
L.Canamero@herts.ac.uk

Abstract

Survival in the animal realm often depends on the ability to elucidate the potentialities for action offered by every situation. This paper argues that affordance learning is a powerful ability for adaptive, embodied, situated agents, and presents a motivation-driven method for their learning. The method proposed considers the agent and its environment as a single unit, thus intrinsically relating agent's interactions to fluctuations of the agent's internal motivation. Being that the motivational state is an expression of the agent's physiology, the existing causality of interactions and their effect on the motivational state is exploited as a principle to learn object affordances. The hypothesis is tested in a Webots 4.0 simulator with a Khepera robot.

1 Introduction

One of the most vital abilities for situated, embedded, autonomous agents in a dynamic scenario is making the right decisions when interacting with their environment. This is the so-called *behaviour or action selection* problem, deciding "what to do next" (what behaviour to execute in a particular situation) to increase the likelihood of maintaining life. Being able to make the right decisions partly depends on the knowledge of the effect of an action to compensate internal needs. Furthermore, it depends on the ability to discriminate objects to befit every interaction. This was confirmed experimentally by Guazzelli et al. (1998), who proposed a behaviour selection model to simulate the behaviour of rats navigating a T-maze, integrating drives and affordances. No perception-related learning was however involved, being that this was solely aimed at interpreting the possibility of moving in one or another direction.

The use of motivational states to make decisions has been proposed in several architectures (Avila-García and Cañamero, 2002; Cañamero, 1997), which mention the necessity not only of maintaining life, but also of meeting the criterion of internal physiological stability (Ashby, 1965). Nevertheless, these architectures neglect the apprehension of the appropriate functionalities of objects. Information about the objects' potential for action has therefore usually been hard-wired. It is argued that knowing the functionality of an object is also part of the adaptation problem.

Related to this, Gibson introduced the notion of *affordance* (Gibson, 1966), defined as the functionality an object offers to an agent. Hence, a set of affordances is only defined in the context of a particular agent-environment pair. Furthermore, *affordances are held to be directly available from the environment, without the integration of perceived features into object representations* (Cooper and Glasspool, 2002). Based on this, Cooper and Glasspool (2002) introduced a symbolic model of affordance learning by relating object features to action schemas. In their approach, object features are symbolically integrated into objects to bias one action or another.

Conversely, the architecture introduced in this paper aims at endowing the agent with the capability of building its own functional perception via an appropriate neural representation of the objects in its environment, related to the agent's behaviour repertoire[1]. This aims at bypassing the feature-based step, and should therefore be a more faithful implementation of gibsonian affordances. Importantly, to perform an action the perception of certain regularities of each object is fundamental to decide the right be-

[1]Unlike Gibson's studies of the optical flow, we have to deal with other perceptual modalities (the agent's senses).

haviour. However, this does not relate to physical resemblance only (among different objects of the same sort), but also to a functional similarity (being able to perform the same actions).

The next section introduces the affordance learning and behaviour selection model, and precedes the experimental section. The paper concludes with a discussion of results and of future research issues.

2 Motivational Model for Learning Affordances

The model comprises three parts: a neural structure, a behaviour arbitration mechanism and a learning module.

2.1 Neural Structure

The first challenge is to build a neural representation of the objects of the environment. To this end, the use of a Growing When Required (GWR) network (Marsland et al., 2002) has been selected. This is a topological network that adapts to the level of entropy of the environment according to a set of parameters, unlike Kohonen (1982). The growing process is described in the following steps:

1. The network is trained with 64-D image patterns representing objects in the scenario. The algorithm chooses the first and second most similar nodes.

2. If the Euclidean distance between the closest node and the current interaction pattern is larger than the pre-set accuracy, a new node is inserted between the two closest nodes, which are then connected by new synapses. Conversely, the closest and its adjacent nodes are dragged towards the input pattern.

3. Nodes rarely close to the patterns are deleted.

4. The growing process is hindered when the euclidean distance between the sensory-patterns and their closest node is smaller than the pre-set level of accuracy.

In a very simple manner, the GWR provides a simple representation of similar objects. The next subsection explains how to relate these patterns to the behaviour repertoire.

2.2 Motivations for Behaviour Selection

The combination of extenal and internal stimuli gives rise to the motivational state. This section describes the necessary elements to build an internal physiology.

The controlled homeostatic variables are abstractions representing an agents' resources. Nutrition, stamina and restlessness are the chosen variables. Their values must be kept within the *viability zone* for the agent to remain alive; if their values overflow/underflow the upper/lower boundaries that define the variable's viability, the robot dies.

The drives are also abstractions denoting urges for action. The drives monitor the levels of the homeostatic variables and initiate a process of compensation whenever they are in a deficit state. In our case, the mechanism of compensation is the selection and execution of a behaviour, which requires an appropriate object nearby for successful execution. We have used three different drives: hunger (which controls nutrition), fatigue (controlling stamina), and curiosity (controlling restlessness). At each time step, each drive is assigned an intensity proportional to the magnitude of the error of its controlled variable.

The behaviours are to grasp, to shelter and to interact. The execution of a behaviour results in an interaction with an object in the environment that may cause a compensation of the deficit for the most critical internal variable, contributing therefore to compensate the drives. In the general case, different behaviours can contribute to compensate a drive, but in our simplified model each drive can be satisfied by one behaviour only, "eat" (grasping an object) satisfies hunger, "shelter" satisfies fatigue, and "interact" satisfies curiosity.

The arbitration mechanism for behaviour selection follows a winner-take-all policy, using the drive that exhibits the highest urgency (the one with the highest level) to choose the behaviour to execute next. In our simplified model this is very straighforward because there is a single behaviour that can satisfy each drive.

The model also has two *Hormones*: Frustration and Satisfaction, which are respectively triggered when the outcome of an interaction succeeds or fails. The values of the hormones are 1, if they are active, and 0 otherwise.

2.3 The Learning Mechanism

The learning process adds a novel dimension to the topological network, by growing *functional* synapses between every node in the aforementioned neural structure and each behaviour of the agent. The procedure for growing these synapses is driven by the agent's drives in a hebbian manner. The process is as follows:

- Every time the agent detects an object, the closest node in the state space is identified. Figure 1 shows the 2D projections of topologies representing the objects contained in the Khepera world used for simulation.
- The interaction succeeds, the hormone Satisfaction is released, otherwise, the hormone Frustration is released.
- Satisfaction and Frustration, strengthen or weaken, respectively, the synapse relating the active node and the behaviour executed ($\Delta\omega_{ij} = \alpha b_j$). Weights are normalised between -1.0 and 1.0.

The final values quantify the *affordances* relating those particular objects, encoded by the neural structure, to the agent's behaviours.

3 Experiments and Results

The goal of these experiments is to test this learning hypothesis with an artificial agent in an engineered scenario. The affordances of the objects in that scenario are such that little objects afford grasping, large objects afford shelter, and all objects afford interacting. Relative sizes vary between 0.08 and 0.01, the size of the Kephera's gripper is 0.04 and the arena measures 0.5×0.5 units.

3.1 Experimental Procedure

The robot wanders in the aforementioned environment, interacting with objects encountered at random. Everytime an object is encountered, the object is centred, and a snapshot of the object is taken always at the same distance. The single top horizontal line of the object is selected, and reduced to a 64-D illumination vector. This vector is used for building the neural structure[2]. Two 2D-PCA of final structures with 16 and 42 nodes each are shown in figure 1.

Concurrently, the agent's homeostatic variables are initialised to their optimal value, and decay according to equation $\Delta hv_i = \tau$, with $\tau = 10^{-5}$. Their optimal values are 0.8 for nutrition and stamina, and 0.2 for restlessness. Their related drives measure the

[2]With parameters $energy = 0.5$, $epsilon_b = 0.5$, $epsilon_n = 0.006$, a_{max}=50, as described in Marsland et al. (2002).

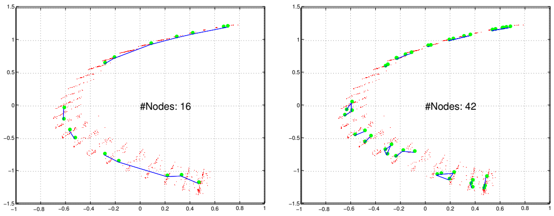

Figure 1: 2D-PCA with GWR overlapping with 16 and 42 nodes, left and right, respectively.

difference from those optimal values, and define the agent's *motivational state*. Whenever an object is encountered, the behaviour whose attached drive exhibits the highest value is selected and executed. The *affordance learning* method, as introduced in section 2.3, is then executed.

3.2 Results

Four series of five simulations each have been run with networks of sizes between 4 and 42 nodes for testing the aformentined learning algorithm. Results for topological networks of 4, 8, 16 and 42 nodes are presented in histograms 2 and 3. The three individual histograms, address the affordance values for each behaviour: grasp, shelter and interact. Values in the X-axis represent the node id in the topological structure, and values in the Y-axis the affordance values learnt (ranged between -1.0 and 1.0), averaged over five simulations.

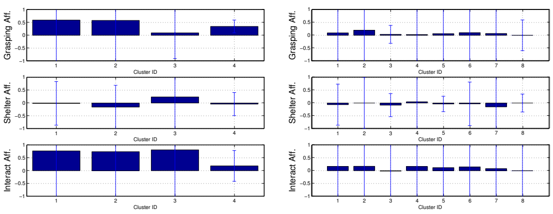

Figure 2: Learnt affordance values for behaviours grasp, shelter and touch (top-down) for GWR with 4 and 8 nodes, left and right, respectively.

It can be observed that affordance values in topologies with a low number of nodes exhibit a large standard deviation. This is due to the low accuracy of those topologies, and is confirmed by observing the difference with affordance values in topologies with a larger number of nodes (16 and 42), which are defined more precisely. In the former case, the low level of accuracy provokes an incorrect selection of the node closest to the visual pattern. In other words,

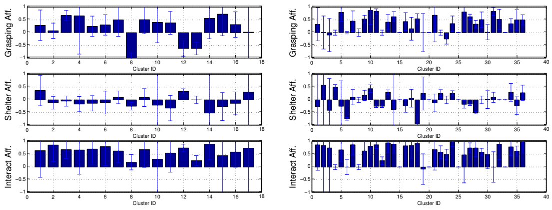

Figure 3: Learnt affordance values for behaviours grasp, shelter and touch (top-down) for GWR with 16 and 32 nodes, left and right, respectively.

nodes in topologies with a low level of accuracy represent a range of objects whose features cannot be causally related to the same effect. This also highlights that for representations of high accuracy, the growing algorithm could be improved via pruning nodes exhibiting affordance values with a high variance. This would not diminish the overall performance, since the resting nodes already represent the sensory-space accurately enough. This would improve the overall performance, since the selection of one node or another would be more accurate, thus its affordance values would be better defined.

Lastly, it is important to highlight that there are implementation and execution issues, e.g., inaccurate object manipulation, which means the execution of some behaviours fail despite the object affording that behaviour to be executed.

4 Conclusions and Future Work

The learning method is based on internal observation of causal fluctuations in the motivational state due to behaviour execution. This provokes a hormonal response, which reinforces the functional synapses relating the behaviour executed to the node in the GWR closest to the perceived sensory pattern. The results suggest that affordances can be learnt according to the experimental procedure proposed.

It is fundamental to stress that affordances are context-related. Hence, to be able *to learn and use affordances, it is necessary to define a context*: the agent's morphology, its set of internal goals and behaviours, the environment. However, sensory perception is independent from the motivational state.

The principles of the model highlight that *motivation and learning are two inter-related processes*. If there is motivation to drive the agent to perform an action, the effect of the performance biases learning. Conversely, learning has a reinforcing role on the motivational (physiological) system. This is grounded in neuroscience by Bindra's suggestion: *"The effects on behaviour produced by reinforcement and motivation arise from a common set of neuro-psychological mechanisms, and the principle of reinforcement is a special case of the more fundamental principle of motivation"* (Bindra, 1969).

Finally, it is relevant to stress that *learning affordances is related to building a representation of the environment; however, a functional representation*. In fact, as the model shows, neural encoding and reinforcement are processes affecting one another.

Future endeavours will perform ethological analysis of behaviour (in terms of physiological stability and cycles of behaviour execution), to assess the effect and reach of this learning process in a variety of environments.

References

W.R. Ashby. *Design for a Brain: The Origin of Adaptive Behaviour*. Chapman & Hall, London, 1965.

Orlando Avila-García and Lola Cañamero. A comparison of behaviour selection architectures using viability indicators. In *Proc. of International Workshop on Biologically-Inspired Robotics: The Legacy of W. Grey Walter*. Bristol HP Labs, UK., August 2002.

D. Bindra. The interrelated mechanisms of reinforcement and motivation, and the nature of their influence on response. In W. J. Arnold and D. Levine, editors, *Nebraska Symposium on Motivation*, pages 1–33. University of Nebraska Press, 1969.

Lola D. Cañamero. Modeling motivations and emotions as a basis for intelligent behavior. In W. Lewis Johnson, editor, *Proceedings of the First International Symposium on Autonomous Agents (Agents'97)*, pages 148–155. New York, NY: ACM Press, 1997.

R. Cooper and D. Glasspool. Learning affordances and action schemas. In R.M. French and J. Sougne, editors, *Connectionist Models of Learning, Development and Evolution*, pages 133–142. Springer-Verlag: London, 2002.

James J. Gibson. *The Senses Considered as Perceptual Systems*. Houghton Mifflin Company, Boston, 1966.

A. Guazzelli, F. J. Corbacho, Bota M., and M.A. Arbib. Affordances, motivation, and the world graph theory. *Adaptive Behavior*, (6(3/4)):435–471, 1998.

Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69, 1982.

Stephen Marsland, Jonathan Shapiro, and Ulrich Nehmzow. A self-organising neural network that grows when required. *Neural Networks*, 15(8-9):1041–1058, 2002.