Long Short-Term Memory for Affordances Learning*

Sergio Roa sergio.roa@dfki.de

Geert-Jan Kruijff gj@dfki.de

German Research Center for Artificial Intelligence / DFKI GmbH

Abstract

This paper addresses the problem of sensorimotor learning from the perspective of affordances learning of simple objects. We are developing a scenario where a robotic arm interacts with a polyflap, a simple 3-dimensional geometrical object. We perform experiments with a simulated arm using a physics simulator, but we plan to use also a real arm. The robot interacts with the object by pushing it in different ways. We use Recurrent Neural Networks to predict the arm and object poses during this interaction, given a discrete set of random actions that the robot can produce.

1. Introduction

Robots should be able to adapt and learn by interacting in dynamic environments, if we want that they acquire the kind of complex skills performed by humans and animals in general. In altricial animals (like humans) the development of complex motor skills is continuously improved after different stages of development. In these species (Sloman and Chappell, 2005), the interaction with the environment plays an important role for the acquisition of sensorimotor abilities, and for the hierarchical acquisition of more complex skills based on the ones previously acquired. This introduces us to the concept of affordance, which is for instance referred to learning about and from actions performed by an agent on an object. In (Gibson, 1977), a theory of affordances was developed. We can apply this theory of cognitive development to the field of robotics by employing, for instance, machine learning techniques that allow the robot to predict action consequences on certain objects. The interaction with objects and in general with different environmental aspects allow to shape the "mind" of the robot on the basis of its acquired experience.

Taking into account that the environment and the physical characteristics (embodiment) of a robot has a complex structure, we have to think of proper scenarios where we can test these techniques and theories. In (Sloman, 2006), simples scenarios using 3-dimensional objects called polyflaps were proposed. The objective is to steadily increase the complexity of the space of actions and the structure of the environment. That would allow us to evaluate algorithms that can be useful for compositional (hierarchical) skills development.

It is also important to identify what kind of perceptions can drive learning for an autonomous robot. Based on the way children acquire learning skills at early stages of development, the works presented in (Oudeyer et al., 2007, Roa et al., 2008) describe a system in which the robot has an intrinsic motivation for learning, based on the interestingness of the situations it discovers. For these tasks, a simple intrinsic reward mechanism is employed, which is proportional to the increase of the error rate of some classifier trying to predict the consequences of the robot actions at a given time. The robot was able to identify *affordances* as correlations between its space and actions and its consequences in the environment. In this work, classifiers are used for prediction and the robot is equipped with real-valued sensors and actions comprising its sensorimotor space. After training, there are different classifiers specialized (biased) in some regions of the state space. A statistical mechanism to split the state space into regions is implemented to support the specialization of the classifiers.

2. Scenario

As already pointed out, we use a robotic arm which interacts with a polyflap in a simulated environment (Figure 1).

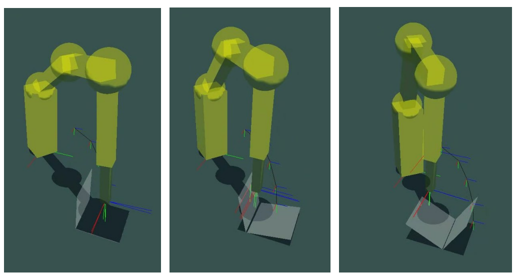

Figure 1: Learning scenario with a polyflap

We use a simulator that can track objects and returns an object pose. Objects that we consider are polyflaps and the arm body parts, which are simple

*The research reported of in this paper is supported by EU FP7 IP "CogX" (ICT-215181)

objects from which we can obtain 3D information. Thus, the task is to use machines that can predict spatio-temporal sequences, and this can be seen as a time-series prediction or regression problem. A sample $\mathbf{s} = [\mathbf{c}, \mathbf{s}_i]_{i=1,\ldots,n}$ is then a whole sequence of feature vectors $\mathbf{s}_i = [\mathbf{v}_i, \mathbf{m}_i]$, where $\mathbf{v}$ denotes a vector containing visual data of an object (pose in homogeneous coordinates), $\mathbf{m}$ denotes motor information (joints pose, joint velocities) and i a time frame number up to the limit $n = 70$, together with a motor control command vector $\mathbf{c}$. In practice, the actions considered are pushing actions on a linear trajectory applying a velocity profile (a 4th degree polynom) to an online inverse kinematics solver and an horizontal direction angle. The values are normalized with mean 0 and standard deviation 1.0.

3. Learning Approach

The learning machines described in (Oudeyer et al., 2007, Roa et al., 2008) can predict short-term consequences of actions. They use an active learning mechanism which uses a measure of learning progress based on the error prediction to select next actions according to this interestingness measure. In this case we are facing a spatio-temporal prediction problem. Recurrent Neural Networks (RNNs), and more specifically Long Short-Term Memory (LSTM) machines (Hochreiter and Schmidhuber, 1997, Graves, 2008) have been shown to accurately predict sequences over extended periods of time. Another approach is the CrySSMEx algorithm(Jacobsson, 2006) which could either extract a probabilistic finite model (a substochastic machine) of the experiences learned by the RNNs (LSTM) or be used itself to analyze the sensorimotor space (as a dynamic system) over several periods of time, and finally extract a model. More importantly, these models should give us a categorization of different object behaviours and corresponding affordances, i.e., given similar objects (similar features) the predictions should be similar. By using these machines, it is possible to evaluate the certainty of the machine to predict action consequences over several periods of time. This mechanism would afford to simulate a kind of mastery driven action selection (if the RNN successfully predicts action consequences) or curiosity driven action selection (if the RNN is failing to predict action consequences and there is learning progress). Other kinds of drives might be novelty (unpredictable action consequence), surprise (unexpected outcome) or interactive (based on a human reward/punishment signal). A feature vector in a frame i is processed at a time step t. The RNN should then predict the corresponding feature vector in the next frame $i+1$ at some time $t+\delta$, till $i = n$. Initially, we use gradient-based methods for offline learning and in online experiments this knowledge might also be used as a kind of knowledge transfer method. In general, a LSTM is composed of input units, special units (gate units, memory cells) or conventional hidden units. The weights w are learned by using a modified gradient descent algorithm, that together with the special units avoid the problem of exponentially decaying error (Hochreiter and Schmidhuber, 1997).

4. Preliminary experimental results

In order to show the convergence of the LSTM machines we performed offline experiments. In a preliminary experiment using 10-fold cross-validation sets and 10 hidden nodes in the network, we obtained the results shown in the experiment 1 in Table 4. SSE denotes the averaged sum of squares error for test sets, which is the objective function minimized by the LSTM and is a good performance measure for regression problems. In the experiment 2, we only used feature vectors $\mathbf{s}_i = \mathbf{v}_i$, i.e., only containing polyflap poses. Because of the non-deterministic nature of a certain control command, slightly different behaviours are produced. We plan to use active learning techniques driven by e.g. curiosity for the selection of samples.

Exp.	Avg. epochs	Avg SSE	Samples
1	4700	0.03	500
2	5622	0.007	500

Table 1: Preliminary results

References

Gibson, J. J. (1977). The theory of affordances. In Shaw, R. and Bransford, J., (Eds.), *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, pages 67–82. Lawrence Erlbaum.

Graves, A. (2008). *Supervised Sequence Labelling with Recurrent Neural Networks*. PhD thesis, Technische Universität München.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, pages 1735–1780.

Jacobsson, H. (2006). The crystallizing substochastic sequential machine extractor - `CrySSMEx`. *Neural Computation*, 18(9):2211–2255.

Oudeyer, P.-Y., Kaplan, F., and Hafner, V. V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 11(1).

Roa, S., Kruijff, G. J., and Jacobsson, H. (2008). Curiosity-driven acquisition of sensorimotor concepts using memory-based active learning. In *Proceedings of the 2008 IEEE International Conference on Robotics and Biomimetics*, pages 665–670.

Sloman, A. (2006). Polyflaps as a domain for perceiving, acting and learning in a 3-D world. In *Position Papers for 2006 AAAI Fellows Symposium*, Menlo Park, CA. AAAI.

Sloman, A. and Chappell, J. (2005). The altricial-precocial spectrum for robots. In *Proceedings IJCAI'05*, pages 1187–1192, Edinburgh. IJCAI.