Active Affordance Learning in Continuous State and Action Spaces

Chang Wang, Koen V. Hindriks and Robert Babuska

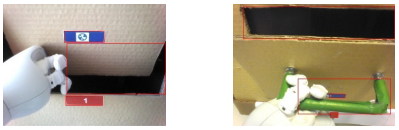
***Abstract*— Learning object affordances and manipulation skills is essential for developing cognitive service robots. We propose an active affordance learning approach in continuous state and action spaces without manual discretization of states or exploratory motor primitives. During exploration in the action space, the robot learns a forward model to predict action effects. It simultaneously updates the active exploration policy through reinforcement learning, whereby the prediction error serves as the intrinsic reward. By using the learned forward model, motor skills are obtained in a bottom-up manner to achieve goal states of an object. We demonstrate that a humanoid robot NAO is able to learn how to manipulate garbage cans with different lids by using different motor skills.**

I. INTRODUCTION

The concept of affordance [3] has been introduced in robotics to address robot-object interaction [6]. Affordances can be modeled as the relations between objects, actions and effects [4]. Learned affordances have been used for predicting action effects and for action planning [8]. However, affordance learning conditions in the literature were strongly controlled by human programmers and this restricts the autonomy of the robot. Not only the spaces of object state and robot action were discretized according to specific tasks, but also the amount of training data was decided before affordance learning actually started. These assumptions do not guarantee that a robot can learn how to manipulate a complex and novel object. In this paper, we take an active affordance learning approach where the robot collects the training data in continuous state and action spaces without manual discretization. We propose active affordance learning in the framework of intrinsically motivated reinforcement learning [1]. Heuristics direct active exploration towards the regions where the prediction errors are maximal [2].

II. AFFORDANCE LEARNING

An *affordance* is defined as the triple: (*Object*, *Action*, *Effect*). We use the manipulation of garbage cans with lids as an example:



(a) Push to open (b) Pull to open

Fig. 1. An illustration of lid manipulation by a robot NAO.

All authors are with the TU Delft Robotics Institute, Delft University of Technology, 2628 CD Delft, The Netherlands. robotics.tudelft.nl. {c.wang-2, k.v.hindriks,r.babuska}@tudelft.nl

The robot identifies an object and the object state based on color segmentation and perceptual proxies such as a bounding box. Affordance learning is to learn an affordance model $\mathcal{F}$ that predicts action effects on an object:

$$e_o = \mathcal{F}(s_o, a, w) \tag{1}$$

where $s_o \in S$ is the state of object o, e.g., lid opening size, $a \in A$ is the action in the 3D Cartesian space, e_o is the effect, e.g., by subtraction of the opening size, and w is the model parameter, e.g., the weight vector of a neural network.

III. ACTIVE AFFORDANCE LEARNING ARCHITECTURE

The overall learning architecture is illustrated in Fig. 2.

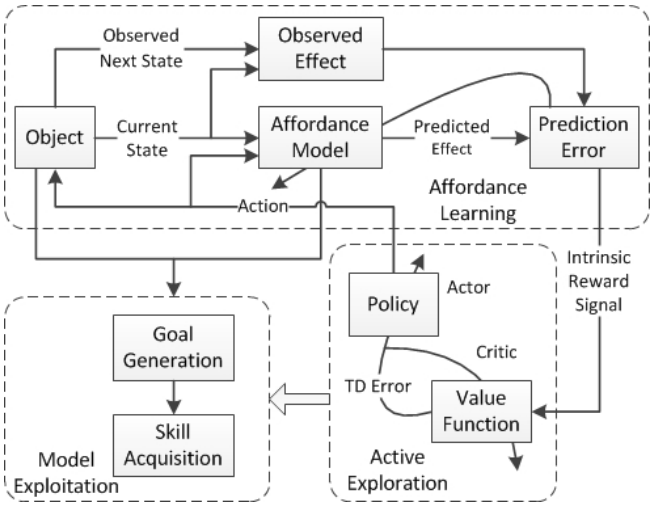


Fig. 2. An architecture of active learning of object affordances.

In the affordance learning component, we use an online version of neural networks to predict action effect: $\hat{e}_o^k = \mathcal{F}(s_o^k, a^k, w^k)$, where k denotes the time step. After the action a^k is applied and the actual next state s_o^{k+1} is observed, the actual effect e_o^k is obtained and the prediction error is calculated: $\eta_k = e_o^k - \hat{e}_o^k$. Then, the new model parameter w^{k+1} is updated: $w^{k+1} = w^k + \alpha\, \eta_k \nabla \mathcal{F}(s_o^k, a^k, w^k)$.

In the active exploration component, we integrate a RL module, i.e., Continuous Actor-Critic Learning Automation (CACLA) [9], in the affordance learning loop. The reward r is generated intrinsically by using the model prediction error: $r = |\eta_k|$. Its maximization is expected to result in an optimal action selection policy. The actor Act_k outputs an action $Act_k(s_o^k)$ based on the current object state s_o^k, and an exploratory action a^k is selected stochastically from the Gaussian probability function $G(x, \mu, \sigma)$ centered around $Act_k(s_o^k)$: $G(x, Act_k(s_o^k), \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x - Act_k(s_o^k))^2/2\sigma^2}$.

The critic V_k learns to predict the value of each state and computes the Temporal Difference (TD) error [7]: $\delta^k = r+\gamma V_k(s_o^{k+1})-V_k(s_o^k)$. The critic is always updated with the TD error: $V_{k+1}(s_o^k) = V_k(s_o^k) + \beta\,\delta^k$. Only if $\delta^k > 0$, the actor Act_k is updated: $Act_{k+1}(s_o^k) = Act_k(s_o^k) + \zeta(a^k - Act_k(s_o^k))$, which means that the performed action a^k is better than expected and should therefore be enforced. The active exploration terminates when the averaged absolute TD error becomes stable, i.e., $|\bar{\delta}^{k+1} - \bar{\delta}^k| < \epsilon$, where ϵ is a small positive threshold, and $\bar{\delta}^k = \frac{1}{N}\Sigma_{i=k-N+1}^{k}|\delta^i|$ is the averaged absolute TD error of recent N actions.

In the model exploitation component, the robot generates goals in the effect space and selects actions to achieve them. For example, the maximized goal effect is: $e_o^g = \arg\max_{a\in A}\,\mathcal{F}(s_o, a, w)$. A range of skills can be acquired in various object states for solving these goal-directed tasks. They are similar to options [5] that are reusable across tasks.

IV. Results

1) Model learning: NAO has learned the near linear relations between object states, robot actions and effects. In the case of the push-lid (see Fig. 1(a)), stretching the arm would result in the opening effect, and stretching further would result in more opening. Besides, the maximal opening effect decreases when the current state of opening increases, and the closing effect was predicted by contracting the arm.

2) Active vs random exploration: The averaged absolute TD error are shown in Fig. 3-4 ($N = 20$). They converged for active exploration while the random exploration failed to converge within allowed number of action steps.

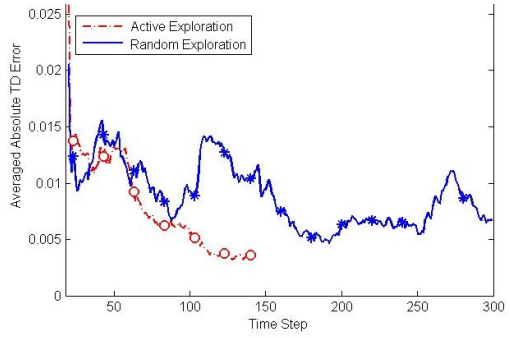


Fig. 3. Experimental result with the push-lid in 3D action space.

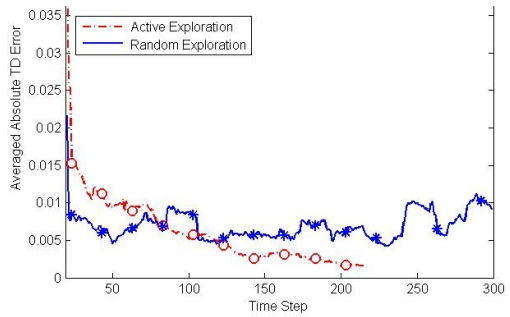


Fig. 4. Experimental result with the pull-lid in 3D action space.

In the active learning mode, NAO intended to explore the most uncertain spaces in an organized way. It usually ended up being blocked by the boundaries of garbage cans, i.e., when a lid was maximally opened or tightly closed. In this case, the object state became stable and no more effect was observed, which gave the TD errors a good chance to converge. In contrast, the random exploration was less efficient because it wasted time on exploring in well predicted action space which contributed little to improving the model prediction accuracy. Besides, it occasionally ran into situations with high prediction errors so that the TD errors would take longer time to converge.

3) Skill acquisition: NAO has learned affordance models to open and close the garbage cans in various object configurations. In the case of opening the push-lid, NAO pushed forward while moving the arm left and up, which resulted in more opening effect observed from NAO's perspective. In the case of opening the pull-lid, NAO pulled the handle while moving the arm downwards. These results agreed with the design of hinges on both lids. A video is available at http://youtu.be/oluLDwMaVoY.

V. Conclusions

We investigated an approach for active affordance learning in continuous state and action spaces for robot use of household products. Affordances were learned on-line to predict action effects meanwhile the prediction error served as intrinsic reward to update the action exploration policy using an actor-critic RL structure. We have demonstrated that a humanoid robot is able to actively learn affordances and efficiently acquire manipulation skills to handle garbage cans. In the future, we will consider the scale of model complexity and the speedup of model convergence, along with the transfer of learned exploration policies for learning novel objects.

References

[1] A. G. Barto. Intrinsic motivation and reinforcement learning. In *Intrinsically Motivated Learning in Natural and Artificial Systems*, pages 17–47. Springer, 2013.

[2] D. A. Cohn, Z. Ghahramani, and M. I. Jordan. Active learning with statistical models. *arXiv preprint cs/9603104*, 1996.

[3] J. J. Gibson. *The ecological approach to visual perception*. Houghton Mifflin, 1979.

[4] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor. Learning object affordances: From sensory–motor coordination to imitation. *IEEE Transactions on Robotics*, 24(1):15 –26, feb. 2008.

[5] T. J. Perkins, D. Precup, et al. Using options for knowledge transfer in reinforcement learning. *University of Massachusetts, Amherst, MA, USA, Tech. Rep*, 1999.

[6] E. Sahin, M. Cakmak, M. R. Dogar, E. Ugur, and G. Ucoluk. To afford or not to afford: A new formalization of affordances toward affordance-based robot control. *Adaptive Behavior*, 15(4):447–472, 2007.

[7] R. Sutton and A. Barto. *Reinforcement learning: An introduction*, volume 1. Cambridge Univ Press, 1998.

[8] E. Ugur, E. Oztop, and E. Sahin. Goal emulation and planning in perceptual space using learned affordances. *Robotics and Autonomous Systems*, 59(7–8):580–595, 2011.

[9] H. van Hasselt and M. A. Wiering. Using continuous action spaces to solve discrete problems. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1149–1156. IEEE, 2009.