# Action Learning and Grounding in Simulated Human-Robot Interactions (Extended Abstract)*

Oliver Roesler and Ann Nowé
Artificial Intelligence Lab, Vrije Universiteit Brussel, Brussels, Belgium
oliver@roesler.co.uk, ann.nowe@vub.ac.be

## 1 Introduction

Service robots that are employed in human-centered environments with a high degree of complexity, unpredictability, and dynamicity must be able to learn new tasks autonomously, i.e. when only the goal of the task is given. Furthermore, robots must be able to understand natural language instructions to accurately identify the requested tasks, which requires connections between symbols, i.e. words, and their meanings, i.e. percepts. There exist many studies in the literature that investigate action learning or grounding, but few consider both simultaneously. Additionally, action learning studies have been limited to learn a single action while only varying the initial position of the gripper [3,4]. Furthermore, grounding studies were mostly conducted offline and primarily focused on grounding of object characteristics or spatial concepts [2,1], while conducted action grounding employed simple feature vectors, which cannot be directly translated into motor commands [5].
In this paper, we investigate the possibility of simultaneous action learning and grounding through the combination of reinforcement and cross-situational learning. More specifically, we simulate human-robot interactions during which a human tutor provides instructions and illustrations of the goal states of the corresponding actions. The robot then learns to reach the desired goals taking into account different manipulation behaviors for different object shapes and grounds the words and detected phrases of the instructions, including synonyms, through obtained percepts.

## 2 System Overview

The employed grounding and action learning system consists of three parts: (1) Human-robot interaction simulation, which generates different situations consisting of the initial gripper and object positions, relative goal positions of the manipulation objects, object colors, object shapes, and natural language instructions, (2) Reinforcement learning algorithm, which employs Q-learning to learn optimal micro-action patterns for encountered situations taking into account initial gripper and object positions as well as the relative goal position of the manipulation object, (3) Cross-situational learning component, which identifies auxiliary words and phrases, and maps percepts to non-auxiliary words and phrases in an unsupervised manner by analysing co-occurrences.

* This is an extended abstract of Roesler and Nowé [6].

## 3 Results

After about 60,000 situations the reinforcement learner required 1 episode to converge to the optimal policy, when using a continuously decreasing exploration rate that is shared across situations. In contrast, when the exploration rate was reset for each situation, the reinforcement learner required after about 9,000 situations on average 28 episodes. That the agent did not execute the optimal policy immediately in the latter case, is due to the high exploration rate at the beginning of each situation because it was reset. Thus, a continuously decreasing exploration rate that is shared across situations works best for the investigated scenario. The employed CSL algorithm is able to successfully ground all 39 words used in this study through their corresponding percepts after about 800 situations. Afterwards the number of correct mappings is constantly 39, while the number of false mappings oscillates between 0 and 2 because the algorithm allows a word to be grounded through several percepts to be able to learn homonyms. The two additional incorrect mappings are for different word combinations, depending on the most recently encountered situations.

## 4 Conclusion

The proposed framework allowed learning of actions through reinforcement learning as well as identification of auxiliary words and phrases, and grounding of words and phrases, including synonyms, through cross-situational learning during simulated human-robot interactions. In future work, the framework will be extended to handle real shape, color and preposition percepts obtained with a stereo camera.

## References

1. Aly, A., Taniguchi, A., Taniguchi, T.: A generative framework for multimodal learning of spatial concepts and object categories: An unsupervised part-of-speech tagging and 3D visual perception based approach. In: IEEE International Conference on Development and Learning and the International Conference on Epigenetic Robotics (ICDL-EpiRob). Lisbon, Portugal (September 2017)
2. Fontanari, J.F., Tikhanoff, V., Cangelosi, A., Ilin, R., Perlovsky, L.I.: Cross-situational learning of object-word mapping using neural modeling fields. Neural Networks **22**(5-6), 579–585 (July-August 2009)
3. Gudimella, A., Story, R., Shaker, M., Kong, R., Brown, M., Shnayder, V., Campos, M.: Deep reinforcement learning for dexterous manipulation with concept networks. CoRR (2017), http://arxiv.org/abs/1709.06977
4. Popov, I., Heess, N., Lillicrap, T., Hafner, R., Barth-Maron, G., Vecerik, M., Lampe, T., Tassa, Y., Erez, T., Riedmiller, M.: Data-efficient deep reinforcement learning for dexterous manipulation. CoRR (2017), http://arxiv.org/abs/1704.03073
5. Roesler, O., Aly, A., Taniguchi, T., Hayashi, Y.: Evaluation of word representations in grounding natural language instructions through computational human-robot interaction. In: Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI). Daegu, South Korea (March 2019)
6. Roesler, O., Nowé, A.: Action learning and grounding in simulated human robot interactions. The Knowledge Engineering Review. (In Press)