Teaching a Robot how to Guide Attention in Child-Robot Learning Interactions

Jacqueline Hemminghaus & Laura Hoffmann & Stefan Kopp
{jhemming, lahoffmann, skopp}@techfak.uni-bielefeld.de
Social Cognitive Systems, CITEC, Bielefeld University, Germany

In this abstract we present ongoing work towards an adaptive social robot that learns natural interaction patterns from child-robot interactions. Therefore, the robot is enabled to interpret children's behavior and react to it appropriately. The major aim of the robot is to support the development of social and communicative skills of children in an educative scenario, where one child is playing a game with the robot. Inter-individual differences in children's developmental stages constitute a major challenge when developing robots for children. Hence, it is difficult to predetermine the behavior of a robot for a child-robot learning interaction. To overcome this problem, our goal is to develop a social robot that is able to learn autonomously, in and from a task-oriented interaction with children, how to employ social behaviors to achieve interactional goals.

We already demonstrated the successful implementation of a social behavior generation for a robot in an adaptive manner [1]. The robot's behavior generation was represented as a hierarchical structure, mapping high-level behaviors, e.g., interactional functions like maintaining attention, onto low-level behaviors that are directly executable by the robot. These low-level behaviors are learned with reinforcement learning [2], enabling the robot to decide by itself how a given high-level behavior is realized through low-level behaviors. This model was evaluated in a human-robot interaction, where the human has to solve a matched pairs memory game displayed on a touch display between the user and the robot. The robot autonomously decided to assist the user to find the matching card by selecting appropriate low-level behaviors to fulfill the goal. The analyses revealed that the learned assistive behavior resulted in a higher task performance compared to a random behavior selection. Additionally, the robot developed different policies according to the user's preferences (e.g., whether they preferred speech or gaze). However, the robot's learned behavior policy did not converge into an optimal solution due to the restricted interaction the participants had with the robot.

In order to handle limited interaction time and thus few data sets for training the robot's behavior policy, we plan to additionally gather demonstrations from the human interaction partner. Therefore, we focused on a different interaction scenario, in which the robot and the human have to fulfill a similar task (here: guide the other's attention). These demonstrations will be used to initialize the robot's behavior policy so that it already has an idea about successful behaviors [3]. In contrast to other learning from demonstration approaches, where the demonstrations are gathered before the actual learning process starts [4], [5], we will gather the demonstrations during the training interaction, where the robot learns its behaviors through experience and refines its policy through the observed behavior of the partner. Demonstrations will be gathered as state-action pairs, where the state describes the current status of the interaction and the action the gestures of the demonstrator.

Study design
With the aim of testing the previously developed model with children and to include demonstrations of humans's behavior in the training phase, we are planning to first collect data from child-robot interactions and afterwards compare the quality of the different extracted policies during a second interaction with other children.

Scenario: As training and evaluation scenario, we choose a simple cooperative game suitable for two players and understandable by children, where both players have the same role. The game will be displayed on a touch display that is placed between the robot and the child. For the interaction we use a sorting game, in which the players have to sort different animals into the correct places, e.g., an ice bear lives in icy regions and not in the savanna. In each round one of the players, Player A, has to select one animal and has to inform the other player, Player B, about its choice without touching the display. In order to confirm what Player B thinks the choice of Player A is, Player B selects the animal on the screen. If its correct, Player A is allowed to move the animal to its right place. In the next turn both players switch their roles, starting with Player B selecting a new animal and so on. One player is played by the robot. In this case it is able to highlight the chosen animal for confirmation through a communication with the touch display. In order to move the

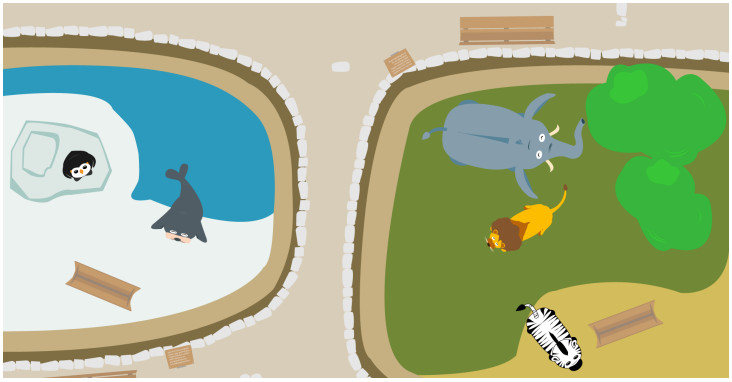

Figure 1: Illustration of the game scenario. Here all animals are already sorted into their right places, indicated by the different compounds.

animal to its right place, the child is asked to do this for the robot, because the robot is not able to touch the screen.

Training Phase: Due to the aforementioned shortcomings of the behavior generation model, the robot learns from reward as before and in addition observes the behavior of the child. When it is the child's turn to chose an animal, the robot observes how the child communicates its choice. This information combined with the current interaction status, e.g., where is the other player looking at and how difficult is the identification of the animal on the game field, are then used to update and improve the robot's behavior policy. The child's attention guiding behaviors are assumed to be always optimal and correct. When it is the robot's turn, it observes the current interaction state and selects either an behavior that was already successful or tries a new behavior in order to explore its possibilities to guide the child's attention to the new selected animal. If the child selects the correct animal the robot will be rewarded for its behavior, otherwise it is punished and needs to clarify its selection. Training is finished, when they cleared the game field together assigning all animals to its correct living areas.

Training data will be collected in two separate conditions. In one condition, the robot learns its behavior policy only with its own experiences while playing the game with the child (comparable to the approach followed in [1]). In the second condition, the robot takes the demonstrations as well as its own experiences to develop its behavior policy into account. The differentiation into these conditions, enables a comparison of the quality of the combined training policy towards the standard reinforcement learning approach. We expect, that the standard combined with online demonstrations lead to a higher task performance than the standard learning approach. Here, task performance will be measured as the reaction time children need to identify the right animal (i.e., touch it), and the amount of errors they produce until they select the right animal.

Testing Phase: To test the success of the previously learned policies, an evaluation study with children will be conducted, in which the learned policies from the training are used and tested with other children. The difference to the training phase will be that the robot will behave according to the previously learned policies. Because during training several policies for one condition are collected, one policy for each condition needs to be extracted. This generalized policy is then used as initialization policy, which could be refined during the testing phase with respect to the child's preferences. In addition to the already collected policies from the training phase, a third policy will be generated from child demonstrations only to compare the impact of this policy to the others. Thus, three conditions will be regarded in the evaluation study: (1) Behavior policy generated by standard reinforcement learning; (2) Behavior policy generated by standard reinforcement learning combined with online demonstrations; (3) Behavior policy generated only from demonstrations.

The three behavior policies will be evaluated according to their quality again. In addition to reaction time and the amount of errors done by the child, the overall acceptance of the robot (e.g., liking, enjoyment of the interaction, intention to interact again with the robot) and the learning success of the children after the interaction will be considered. Therefore, children will be tested about their knowledge of animals and their habitats before and after the interaction with the robot. The results of the study will be presented at the workshop.

References

[1] J. Hemminghaus and S. Kopp, "Towards adaptive social behavior generation for assistive robots using reinforcement learning", in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ACM, 2017, pp. 332–340.

[2] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, 1. MIT press Cambridge, 1998, vol. 1.

[3] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration", *Robotics and autonomous systems*, vol. 57, no. 5, pp. 469–483, 2009.

[4] B. Kim, A. massoud Farahmand, J. Pineau, and D. Precup, "Learning from limited demonstrations", in *Advances in Neural Information Processing Systems*, 2013, pp. 2859–2867.

[5] T. Brys, A. Harutyunyan, H. B. Suay, S. Chernova, M. E. Taylor, and A. Nowé, "Reinforcement learning from demonstration through shaping.", in *IJCAI*, 2015, pp. 3352–3358.