ROBOT LEARNING FROM FEEDBACK

Joost Broekens
Delft University of Technology
Delft
The Netherlands
Joost.broekens@gmail.com

Synonyms

Interactive reinforcement learning, shaping.

Definition

In the context of robot learning via human-robot interactions, robot learning from feedback refers to the ability of a robot to change future behavior based on feedback given by a human during the learning process. Consequently, feedback in this case refers to a signal from a human observer indicating the appropriateness of a particular behavior of the robot in a particular spatiotemporal context.

Theoretical Background

Robots have been used for a long time to help humans cope with repetitive or hazardous tasks. Robots either take over these tasks and function autonomously, or are used as "eyes and hands" through teleoperation, e.g., in manufacturing or search and rescue respectively. However, robots are more and more studied in social and service oriented settings (Fong, Nourbakhsh, & Dautenhahn, 2003), such as elderly care and work at service or information desks.

Typically, such social robots need to interact with people in order to adapt their functionality. Human robot interaction thus becomes an important issue. One of the main reasons for investigating natural ways of interacting with robots in order for the robot to adapt its behavior is *non-expert interaction;* i.e., to enable persons not familiar with robot learning to change the behavior of robots in a natural way (Thomaz & Breazeal, 2006).

Here we focus on robot learning by human feedback, but we briefly discuss two other approaches in this theoretical background section. In general there are three types of robot teaching (Thomaz & Breazeal, 2006): by *example*, by *guidance* and *feedback*.

In the case of learning by *example*, robots learn behavior by imitating human behavior when that behavior is provided as an example. The robot either imitates the behavior, or imitates getting towards the goal using the behavior as example. In the first case the behavior is leading, in the second the intention of the behavior is leading.

In general, robot *guidance* is about directing attention, communicating motivational intentions, and proposing actions (Thomaz & Breazeal, 2006). Learning by guidance can be differentiated from feedback and imitation in the following way. While feedback gives intentional information after the fact, guidance gives intentional information before the fact (Thomaz & Breazeal, 2006). For example, smiling at a robot after it has taken the right turn towards the food (feedback) is quite different from proposing a certain turn to the robot before it has chosen itself (guidance). While imitation refers to the repetition of a sequence of actions that lead towards a goal state, guidance is about future-directed learning cues and as such is much broader defined. For example, showing a child how to tie shoelaces (example) is very different from drawing a child's attention to the two edges when stuck in the beginning (guidance).

Learning from *feedback* refers to using a human signal as information about the appropriateness of past actions (behavior). In this case, the signal is used "after the fact" to adapt the behavioral strategies of the robot. Usually, such robots use reinforcement learning (Broekens, 2007, Thomaz & Breazeal, 2006, Thrun et al, 1999, Knox & Stone, 2010) as a basis. Reinforcement learning (RL) is based on the idea that an agent learns behavior by exploring an environment (state space) and learning which actions to repeat and avoid based on positive and negative feedback respectively. This is compatible with instrumental condition in psychology. RL has been used extensively and in a wide variety of contexts. Importantly, RL assumes that the feedback consists of a reward or punishment that is used for learning as follows: the reward is propagated back over the sequences of actions responsible for the reward in such a way that the values attached to these actions converge to the expected cumulative future reward. In short, the reward or punishment is a signal "after the fact" one could refer to as *pure feedback*. This pure feedback is used to change the values of past actions.

Important Scientific Research and Open Questions

In general there are three main (and currently unresolved) challenges to address. First, *detecting and interpreting* a signal from a user as feedback. Second, *using this feedback* in the context of a robot learning mechanism. Third, doing this in *real-time*. We now detail these challenges.

Feedback *detection* means that behavior of the user (interaction partner) of a robot is interpreted as meaningful to the learning process or not. Meaningful in this case can be specifically interpreted as reinforcing, or disapproving of, a particular behavior. For example, a user can influence the robot's learning by expressing affective expressions (Broekens, 2007), where positive expressions are interpreted as reward and negative expressions as punishment. In another setting, the feedback can be generated based on the presence of (the number of) humans, not a specific signal from a user. For example, Thrun (1999) show how a museum tour guiding robot learns to optimize tour-guiding behavior by using the number of people present as learning signal. When trying to detect the feedback signal from a human, there is problem: humans give feedback in a mixed way (Kim et al ,2009). Types of feedback humans give to learning robots include *pure feedback*, *anticipatory feedback*, *attention guidance* and mixtures of these. This is an extremely important observation as it means that human feedback cannot simply be equated with reward in RL (Kim et al., 2009; Thomaz & Breazeal, 2006, Knox & Stone, 2010), and hence feedback detection becomes relevant, as these different kind of signals mean completely different things from a learning perspective. A pure reward can be interpreted as reward in RL, but an anticipatory should be interpreted differently, for example as guiding attention by influencing action selection. So, there are two major issues in feedback detection: the *form* of the signal (what is the feedback signal and how does the user communicate this), and the *intention* of the signal (what does it mean in the context of the robot's past and future behavior).

Using the feedback means that once a signal is detected as a reward, it needs to be embedded in the robot's learning mechanism. For example, when a cleaning robot just broke a vase, a user expresses anger, and the expression is interpreted as feedback about the robot's behavior. Now, the question is how to use the feedback in the learning mechanism so that the robot adapts its future behavior. Knox and Stone (2010) study a large variety of different interpretations of a human feedback signal in the context of robot learning based on reinforcement learning. These varieties include:

- feedback as pure reward added to the reward function R of the environment ($R=R_{environment}+R_{human)}$
- feedback as target for the value function V to learn (approximate V using R_{human}), and
- feedback as value that is added to the to-be-learned value function ($V=V+R_{human}$).

Note that the last two varieties already imply that the reward is not a pure reward, as the value function in RL expresses the accumulation of future to-be-expected reward, so in essence the feedback is already being used as an anticipatory signal. Currently it is not clear what the best ways are to embed human feedback in reinforcement learning (and related) mechanisms.

Finally, doing the previous two processes in *real-time* is a problem for two main reasons. First, detection of feedback signals is far from trivial and involves understanding the context of the user and robot (what did the robot just do, what does my user like/dislike), a model of interpretation of the signal (explained above) and attention detection mechanisms (is the feedback for the robot). Second, reinforcement learning typically assumes a separate learning and a performing phase. In the learning phase, the robot learns to shape its behavior, in the performing phase, the robot uses what it has learned. However, humans keep giving feedback, so there is no separation between these phases making it an *online* learning problem. Therefore, robot learning mechanisms that use human feedback must be able to continuously integrate human feedback.

Cross-References

→ Reinforcement learning

→ Robot learning

→ Feedback and learning

→ Computational models of conditioning

→ Robot learning via human-robot interactions

→ Robot learning from demonstrations

→ Imitation learning of robots

→ Imitation and learning

References

Broekens, J. (2007). Emotion and Reinforcement: Affective Facial Expressions Facilitate Robot Learning. In Artifical Intelligence for Human Computing (pp. 113-132).

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. Robotics and Autonomous Systems, 42(3-4), 143-166.

Kim, E. S., Leyzberg, D., Tsui, K. M., & Scassellati, B. (2009). How people talk when teaching a robot. Paper presented at the Proceedings of the 4th ACM/IEEE international conference on Human robot interaction.

Knox, W. B. & Stone, P. (2010). Combining manual feedback with subsequent MDP reward signals for reinforcement learning. In: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010) (pp. 5-12).

Thomaz, A.L., Breazeal, C.: Teachable characters: User studies, design principles, and learning performance. In: Gratch, J., Young, M., Aylett, R., Ballin, D., Olivier, P. (eds.): Proc. of the 6th International Conference on Intelligent Virtual Agents (IVA 2006). LNCS, Vol. 4133, Springer (2006) 395-406

Thrun, S., Bennewitz, M., Burgard, W., Cremers, A., Dellaert, F., Fox, D., et al. (1999). MINERVA: A Tour-Guide Robot that Learns. In KI-99: Advances in Artificial Intelligence (pp. 696-696).