A posteriori response distinction versus apriori response prediction

Inna Mikhailova   **Liyan Zhang**   **Christian Goerick**
Honda Research Institute Europe GmbH,
Carl-Legien-Strasse 30,
63073 Offenbach / Main, Germany
inna.mikhailova@honda-ri.de

One of the most intriguing questions in developmental robotics is how to bootstrap the learning process. In our work, we design a system that is potentially able to learn in highly dynamical situations like the interaction with a human.

There is a crucial difference between the learning of the response of an object and the learning of the response of a human. Although humans give typical responses, these responses can strongly vary depending on the hidden state of the human. Mostly the state can be detected only afterwards and no reliable prediction can be done. To our knowledge, until now, the implementations that use the increasing predictability of an action response as a drive of the developmental process (Oudeyer and Kaplan, 2004), (Singh et al., 2004) do not cover the possibility of actions with multimodal response distributions depending on hidden states of the interaction partner. However, such actions are an important part of the behavioural repertoire. They can be used to enquire those hidden states.

A test scenario for our framework is the autonomous learning of a request gesture. The reflexive bootstrapping of robot-human interaction consists of the abilities to track an object with the cameras and to stretch out the hand towards the object. The robot explores the effect of changing its palm orientation. The hand outstretched with the palm pointing upwards is typically interpreted as a request (Fig 1, b)). For other orientations of the palm the ambiguity of the user response is higher because the user can interpret even a showing gesture as a request. The response is therefore not fully predictable because it depends on the non-observable interpretation of the robot's gesture by the human.

The robot acquires possible classes of user responses (Fig. 1, a)) using an online learning of a Gaussian Mixture Model (GMM). The user response is measured as a velocity of the object between two subsequent stops. If the user gives the object to the robot, it is assumed that the object has a positive velocity. In this way we obviate the definition of a particular time to measure the user's reaction, thus avoiding the problems described in (Doniec et al., 2006). Only a time-out while waiting for the reaction needs to be specified.

The GMM is used to make predictions and to control the hand orientation. We implicitly encoded a drive towards unambiguous actions response ("controllability") in the action selection mechanism. The action is chosen such as to maintain the user response with the highest probability. Due to the normalisation factor in the conditioned distribution $P(r|a)$ of the response r given action a ($P(r|a) = P(a, r) / \int P(a, r) dr$) the probability of getting the response is higher for unambiguous actions. This property leads to a natural exploration of the action parameters towards higher controllability.

As we accept a multimodal response distribution, the prediction error does not lead to a change of the model if there exists a cluster that explains the observed data well. We will discuss next how we deal with the discrepancy between the a-priori prediction based on the long-term statistics and the a-posteriori classification of a current situation.

The "classical" GMM makes the assumption of observing a stationary process. The weights of the mixture describe the statistics of the cluster activations over a very long observation time. However, the activation of a cluster can be context-dependent and thus can have different statistics on a short time-scale than on a long time-scale. Therefore we need both short-term and long-term statistics for better control. In the described scenario, the hidden state (cluster activation) corresponds to the user's interpretation of the robot's gesture. We capture the lower ambiguity of the "gimme" gesture in a probability distribution calculated over a long time-scale (Fig. 1, a)). In contrast, we capture how a particular user reacts to the gesture in a current context by a distribution of the a-posteriori cluster weights calculated over a short time-scale. For prediction we combine the long- and short-term statistics in the GMM. We use the a-priori weights for the clusters that do not overlap with the currently executed gesture, while using the short-term statistics of a-
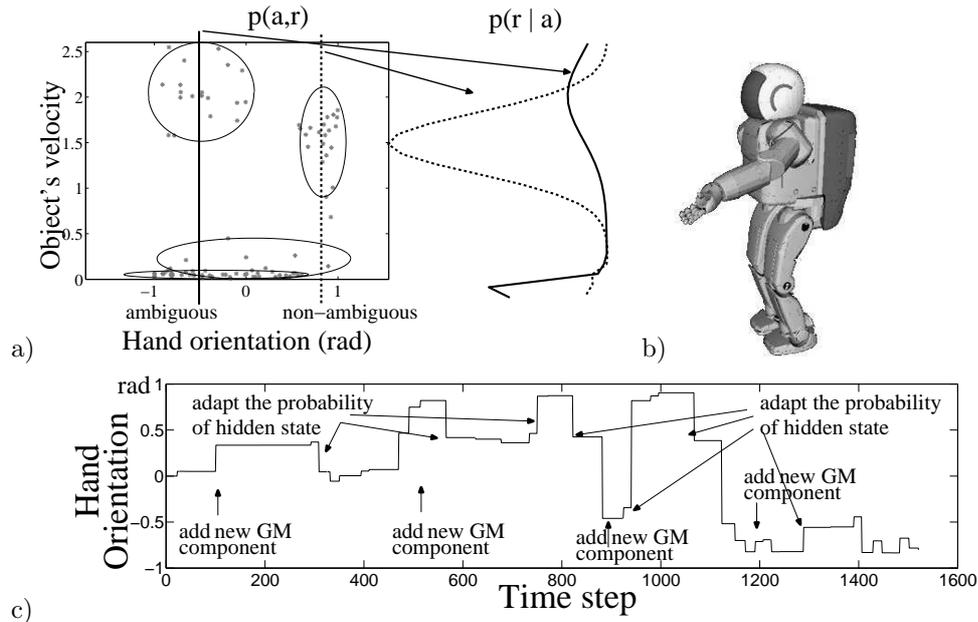
Figure 1: Autonomous acquisition of request gesture. a)The learnt action-response distribution. The solid line shows the probability of the response for the ambiguous the dashed line for the non-ambiguous gesture. b)The non-ambiguous request gesture with the palm turned up. c)Typical run of online-learning and control of the hand orientation, one time step is about 0.25 seconds.

posteriori weights for clusters that includes the executed gesture. In this way, the accumulation of the short-term statistics serves as an enquiry about the hidden state (Fig. 1, c)).

In the future, a prediction error on the level of hidden state activation (difference between the a-priori and a-posteriori GMM weights) can be used for the learning on higher abstraction levels (in similar terms as described by (Wolpert and Kawato, 1998)). For example, the system may learn the correlation of hidden states to observable variables. These could be the user's speech that indicates the interpretation of a gesture or a decision not to give an object, as well as observable properties of the objects that can or cannot be given. If the system learns this correlation, then it can switch faster between states. Without additional evidence for a state switch we need a certain time for the accumulation of the a-posteriori statistics in order to prevent an undesired reaction to outliers.

The presented work stresses the necessity of differentiated dealing with prediction errors. On the one hand, these errors can mean that the system has to change the model on the level of the observed data. On the other hand, they may be indicators for the need of differentiation and learning on a higher abstraction level. Finally, for some types of interaction with the environment prediction errors may be not avoidable, e.g., in interaction with a human. In these cases it may be more appropriate to use a multimodal response distribution and an a-posteriori

response differentiation as a source of information for further acting. Therefore, the developmental drive is not to make no prediction errors but to make informative errors while keeping the cost of error making as small as possible. The progress of the system may be expressed by a better choice of enquiry about the hidden states. "Better" can mean less energy consuming and more informative. E.g., in our scenario, the system should adapt its action from a fully outstretched hand towards a more relaxed gesture accompanied by speech.

References

Doniec, M. W., Sun, G., and Scassellati, B. (2006). Active learning of joint attention. In *IEEE-RAS International Conference on Humanoid Robots (Humanoids 2006), Genova, Italy.*

Oudeyer, P.-Y. and Kaplan, F. (2004). Intelligent adaptive curiosity: a source of self-development. In *Proceedings of the 4th International Workshop on Epigenetic Robotics, Genova, Italy.*

Singh, S. P., Barto, A. G., and Chentanez, N. (2004). Intrinsically motivated reinforcement learning. In *18th Annual Conference on Neural Information Processing Systems (NIPS), Vancouver, B.C., Canada.*

Wolpert, D. and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, 11:1317–1329.