Autonomous Learning Evaluation toward Active Motor Babbling

Ryo Saegusa[1,a], Sophie Sakka[2], Giorgio Metta[1,3], Giulio Sandini[1,3]

[1]Italian Institute of Technology, Genoa, Italy
[2]University of Poitiers, Poitiers, France
[3]University of Genoa, Genoa, Italy
[a]E-mail: ryos@ieee.org

Abstract

Learning in robotics is one of the practical solutions allowing an autonomous robot to perceive its body and the environment. As discussed in the context of the *frame problem* [1], the robot's body and the environment are too complex to be modeled. Even if the kinematics and the dynamics of the body are known, a real sensory input to the body would be different to one derived from the theoretical model, because the sensory input is always influenced by the interaction with the environment. For instance, when we grasp an object, the physical state of our arm such as a weight and momentum becomes different to those at the normal state. However, it is difficult to evaluate all potential variation in advance, since real data can vary quite a lot and the behavior of the external environment is not necessarily controlled by the robot: in this example, the state of the arm is always different depending on the grasped object. On the other hand, learning provides a data-driven solution: the robot explores the environment and extracts knowledge to build an internal model of the body and the environment.

Learning-based motor control system is well studied in the literature [2] [3] [4] [5] [6] [7]. Haruno et al. proposed a modular control approach [3], which couples a forward model (state predictor) and an inverse model (controller). The forward model predicts the next state from a current state and a motor command (an efference copy), while the inverse model generates a motor command from the current state and the predicted state. The desired motor command is not available, but the feedback error learning procedure (FEL) provides a suitable approximation [4]. The prediction error contributes to gate learning of the forward and inverse models, and to weight output of the inverse models for the final motor command. Motor prediction based on a copy of motor command compensates the delays and noise in the sensorimotor system. Moreover, motor prediction allows differentiating self-generated movements from externally imposed forces/disturbances [5][6].

Learning-based perception is applicable not only for motor control but also to model the environment owing to multiple sensorial modalities, such as vision, audition, touch, force/torque, and acceleration sensing. In a similar approach, we developed a learning system aiming at predicting future sensing data based

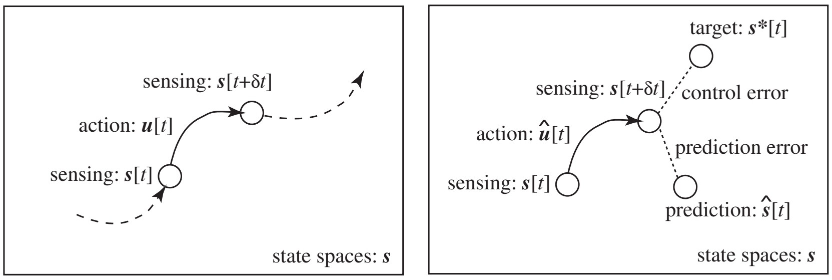

Figure 1: Internal state space. Left: ideal state transision, Right: estimated state transision.

on current sensing data and motor command [8]. Unlike most studies on sensory-motor prediction, the robot and the environment are considered dynamic. Thus, we explored the possibilities for the robot to detect changes in its self or its environment in an autonomous manner: no other information such as a model was given to the system. Following this concept, we investigated a function called *confidence*, driven in the evaluation process of the sensory prediction learning [9]. The aim of this function is to detect inequalities between the predicted situation and the real situation of the body and the environment. The notion of robotic self-confidence was developed as the first step toward self diagnosis and self adaptation.

Our global aim is to implement a learning process as a natural adaptation and self-improvement for the robot. In this context, one of the significant problems in learning is that it requires much time for data sampling and post treatment. An efficient learning strategy is necessary to enhance the learning speed while keeping its quality. The random sampling strategy is considered as the most robust approach for unknown learning environment, on the other hand maybe there are some more formal ways of choosing the sampling strategy depending on various factors and constrains of the body and the environment, which biases robot learning interest. We propose an improvement of the learning strategy (active motor babbling) based on the confidence function with multiple sensory modalities: the evaluation of learning is applied to the next exploration of data sampling.

Fig.1 illustrates internal state space of the robot. Let $\boldsymbol{s}[t] \in R^{n_s}$ denote the sensory state vector for the n_s sensors, and $\boldsymbol{u}[t] \in R^{n_m}$ be the motor command vector for the n_m motors at time t. Let us consider that the dynamics of $\boldsymbol{s}[t]$ and $\boldsymbol{u}[t]$ can be defined as:

$$\boldsymbol{s}[t+\delta t] \quad := \quad \Phi(\boldsymbol{s}[t], \boldsymbol{u}[t]), \tag{1}$$

$$\boldsymbol{u}[t] \quad := \quad \Psi(\boldsymbol{s}[t], \boldsymbol{s}[t+\delta t]). \tag{2}$$

Here, for simplicity, an action to transit the state from $\boldsymbol{s}[t]$ to $\boldsymbol{s}[t+\delta t]$ is assumed as unique. The goal of learning is to approximate $\Phi(\cdot)$ and $\Psi(\cdot)$ using data samples acquired through exploration. Let $\hat{\boldsymbol{s}}[t]$ and $\hat{\boldsymbol{u}}[t]$ denote estimated vectors of the next sensory feedback: $\boldsymbol{s}[t+\delta t]$ and the actuated motor command: $\boldsymbol{u}[t]$. $\hat{\Phi}(\cdot)$ and $\hat{\Psi}(\cdot)$ denote the approximations of $\Phi(\cdot)$ and $\Psi(\cdot)$:

$$\hat{\boldsymbol{s}}[t] \quad := \quad \hat{\Phi}(\boldsymbol{s}[t], \boldsymbol{u}[t]), \tag{3}$$

$$\hat{\boldsymbol{u}}[t] \quad := \quad \hat{\Psi}(\boldsymbol{s}[t], \boldsymbol{s}[t+\delta t]). \tag{4}$$

In order to collect learning data, the robot must explore the environment. However, in the begining of learning, the robot does not know how to explore. A motor babbling behavior gives a simple solution: the robot generates $\boldsymbol{u}[t_k]$ randomly, and stores learning data $\{\boldsymbol{s}[t_k], \boldsymbol{u}[t_k], \boldsymbol{s}[t_{k+1}]\}_{k=1,\cdots,K}$ at each time step. $\hat{\Phi}(\cdot)$ and $\hat{\Psi}(\cdot)$ can be optimized by using acquired data $\boldsymbol{s}[t_k]$ for $\boldsymbol{s}[t]$, $\boldsymbol{u}[t_k]$ for $\boldsymbol{u}[t]$ and $\hat{\boldsymbol{u}}[t]$, and $\boldsymbol{s}[t_{k+1}]$ for $\boldsymbol{s}[t+\delta t]$ and $\hat{\boldsymbol{s}}[t]$.

If the learning is well performed, the robot is able to generate a motor command to reach the desired next state: $\boldsymbol{s}^*[t]$, which means that the robot is able to perform motor babbling not in the joint space: $\boldsymbol{u}$, but in the state space: $\boldsymbol{s}$, defined as:

$$\hat{\boldsymbol{s}}[t] \quad := \quad \hat{\Phi}(\boldsymbol{s}[t], \hat{\boldsymbol{u}}[t]), \tag{5}$$

$$\hat{\boldsymbol{u}}[t] \quad := \quad \hat{\Psi}(\boldsymbol{s}[t], \boldsymbol{s}^*[t]). \tag{6}$$

Learning result can be evaluated as *confidence* for the state. The confidence is based on the state prediction error: $||\boldsymbol{e}_\phi||$ and motor control error: $||\boldsymbol{e}_\psi||$ defined as:

$$\boldsymbol{e}_\phi[t] \quad := \quad \hat{\boldsymbol{s}}[t-\delta t] - \boldsymbol{s}[t], \tag{7}$$

$$\boldsymbol{e}_\psi[t] \quad := \quad \boldsymbol{s}^*[t-\delta t] - \boldsymbol{s}[t], \tag{8}$$

where $\hat{\boldsymbol{s}}[t-\delta t]$ indicates the prediction of $\boldsymbol{s}[t]$ executed at time $t-\delta t$, and $\boldsymbol{s}^*[t-\delta t]$ indicates the target state for control at time $t-\delta t$. The domain of the components of $\boldsymbol{e}_\phi[t]$ and $\boldsymbol{e}_\psi[t]$ has no boundary: $(-\infty, +\infty)$. Here, let us introduce a transformation of $\boldsymbol{e}_\phi[t]$ and $\boldsymbol{e}_\psi[t]$ into a finite scalar variable: $c[t] \in [0,1]$ such as

$$c[t] := \exp(-||\boldsymbol{e}_\phi||^2[t] \ / \ 2\sigma_\phi^2) \cdot \exp(-||\boldsymbol{e}_\psi||^2[t] \ / \ 2\sigma_\psi^2), \tag{9}$$

where the variances σ_ϕ^2 and σ_ψ^2 determine sensitivity. Accumulation of $c[t]$ depending on the sensory state $\boldsymbol{s}[t]$ provides confidence for a state. Let $C[\boldsymbol{s}] \in [0,1]$ denote the confidence: a high value of $C[\boldsymbol{s}]$ means that learning of state dynamics at $\boldsymbol{s}$ is reliable. The update rule of the confidence at time $t+\delta t$ is defined as:

$$C[\boldsymbol{s}, t+\delta t] := (1-\alpha)C[\boldsymbol{s}, t] + \alpha c[t]. \tag{10}$$

where the constant parameter $\alpha \in [0,1]$ is an update weight, and $C[\boldsymbol{s}, 0]$ is initialized as zero. The confidence works as a temporal moving average of normalized learning error.

The principal idea of this framing is to exploit confidence derived from the past learning, for the next exploration to collect new learning data. If the confidence at the current state is low, the robot generates random motor babbling in $\boldsymbol{u}$ space, while the confidence is high, the robot directs its action, in $\boldsymbol{s}$ space, into the lower confidence state to collect new learning data for improvement. In this work, we are going to discuss the behavior of the algorithm resulting from the experimental learning using the humanoid robot James, shown in Fig.2.

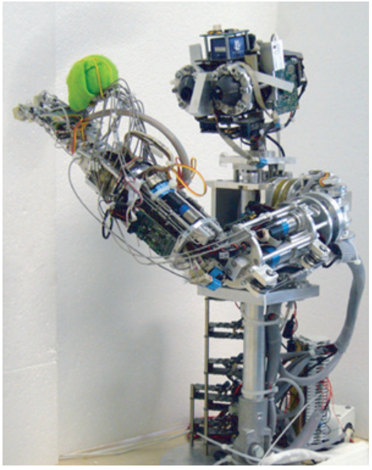

Figure 2: The humanoid robot James [10].

References

[1] J. McCarthy, P. J. Hayes, Some philosophical problems from the standpoint of artificial intelligence, Machine Intelligence 4, pp.463–502, 1969.

[2] M. I. Jordan, D. E. Rumelhart, Forward models: Supervised learning with a distal teacher, Cognitive Science, 16(3), pp.307–354, 1992.

[3] M. Haruno, D. M. Wolpert, M. Kawato, MOSAIC Model for sensorimotor learning and control, Neural Computation, 13, pp.2201–2220, 2001.

[4] M. Kawato, Internal models for motor control and trajectory planning, Current Opinion in Neurobiology, 9, pp. 718–727, 1999.

[5] D. M. Wolpert, Z. Ghahramani, R. J. Flanagan, Perspectives and problems in motor learning, Trends in Cognitive Sciences, 5(11), pp.487–494, 2001.

[6] D. M. Wolpert, R. C. Miall, Forward models for physiological motor control, Neural Networks, 9(8), pp.1265–1279, 1996.

[7] D.M. Wolpert, JR Flanagan, Motor Prediction, Current Biology 11(18) R729-732, 2001

[8] R. Saegusa, F. Nori, G. Sandini, G. Metta, S. Sakka, Sensory prediction for autonomous robots, IEEE-RAS 7th International Conference on Humanoid Robots (Humanoids2007), Pittsburgh, USA, 2007.

[9] R. Saegusa, S. Sakka, G. Metta, G. Sandini, Sensory prediction learning – How to model the self and environment–, The 12th IMEKO TC1-TC7 joint symposium on Man, Science and Measurement, Annecy, France, 2008. (to appear)

[10] L. Jamone, G. Metta, F. Nori and G. Sandini, James, a humanoid robot acting over an unstructured world, Proc. of the Humanoids 2006 conference, pp.143–150, Genoa, Italy, 2006.