A Developmental Approach Accelerates Learning of Joint Attention

Yukie Nagai Minoru Asada Koh Hosoda

Dept. of Adaptive Machine Systems, Graduate School of Engineering,
Osaka University, Suita, Osaka, 565-0871 Japan
e-mail: yukie@er.ams.eng.osaka-u.ac.jp

Abstract

This paper argues how robot learning can be accelerated by a developmental approach that changes the capability of the learning robot (robot development) and/or the complexity of the environment (environmental development). "Robot development" means that the perceptual, cognitive, and behavioral capabilities of the learning robot change from immature to mature states in accordance with learning process. On the other hand, "environmental development" means that the complexity of the environmental structure changes from simple to complicated ones in accordance with learning process. Then, the developmental approaches can be categorized according to the relationships between what does develop and what does make the development. A case study of developmental robot learning for joint attention in which a caregiver can be regarded as a part of the environment is given to show how it can accelerate the learning. Finally, future work and discussion are given.

1 Introduction

Developmental approaches to robot learning have potential to accelerate the learning and/or to improve the final performance of the learned behaviors. It is well known in developmental cognitive science that the developments of attentional and memorial abilities of children help to learn their first language [5, 13]. In the early stage of learning of the language, children are not able to understand all what adults spoke to them, because their abilities of attention and memory are limited. However, the limitation of these abilities become a useful filter to segment and to identify the components that comprise the language, and as a result this filter makes it easier to learn the language. On the other hand, parents, teachers, and other adults adapt themselves to the need of children according to each child's level of maturity and particular relationship they have developed with the child [8]. Thus, the developments would happen in both sides: children and adults.

This can be regarded that two kinds of development structures interact with each other via learning process: the internal development structure inside the learning agent and the environmental development one outside the agent. We may apply this idea to a robot in order to accelerate its learning. The former means that the perceptual, cognitive, and behavioral capabilities of the learning robot change from immature state to mature one with learning process. The latter means that the environmental complexity changes from the lower level to the higher one so that it can accelerate the learning.

In this paper, we argue how the developmental approaches can accelerate robot learning from a view point of cognitive developmental robotics [1]. First, developmental approaches are categorized according to the relationships between what does develop and what does make the development. Next, a joint attention problem between a human and a robot is given to evaluate the developmental approaches. Joint attention is a challenging problem, because it is a fundamental capability for communication and lead to acquisition of other social capabilities, such as language understanding and mind reading [12]. Finally, conclusions and future works toward shared attention of which realization may need a more developmental approach are discussed.

2 Developmental Approach

The developmental approaches to the robot learning are categorized according to the relationships between what does develop (a robot or its environment) and what does make the development (trigger to develop). Therefore, the developmental approaches are categorized as follows,

(A-1) robot development triggered by the given clock,

(A-2) robot development triggered by the learning performance,

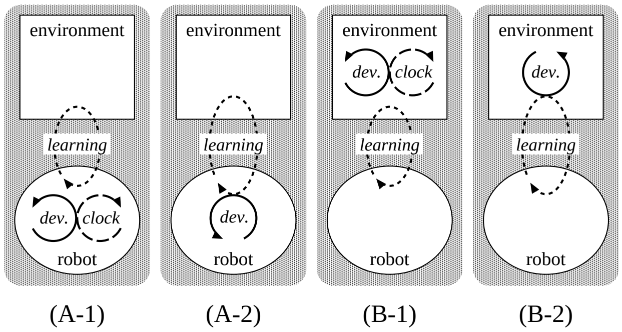


Figure 1: Relationships between developmental agents and its triggers

(B-1) environmental development triggered by the given clock, and

(B-2) environmental development triggered by the learning performance.

Figure 1 shows the relationships between a developmental agent and a trigger. Here, the robot development means that the perceptual, cognitive, and behavioral capabilities of the learning robot change from immature to mature states. On the other hand, the environmental development means that the complexity of the environmental structure that includes a caregiver changes from simple to complicated ones.

2.1 Robot Development Triggered by The Given Clock

There is a developmental approach caused by the internal clock. The clock works periodically, that is independent of the learning process. More specifically, this developmental approach means that the perceptual and behavioral capabilities of the learning robot change from immature to mature states regardless of the learning performance. For example, the resolution of the receptive field turns from coarse to fine ones according to the schedule by the internal clock.

Dominguez and Jacobs [4] showed that the learning of binocular disparity sensitivities by a developmental model with a clock was better than that by a non-developmental model. The developmental model received the lower spatial frequency information during the early stage of learning, and then the higher frequency information was added at the pre-specified time clock. In contrast, the non-developmental model received all kinds of spatial frequency information during the all learning stages. Metta *et al.* [11] showed that learning of saccadic motion of robot's camera was accelerated by the similar developmental approach to Dominguez and Jacobs' one [4]. They constructed their approach by changing the resolution of camera from coarse to fine state.

These approaches of development caused by a clock can accelerate learning or improve the final performance according to the pre-specified time schedule. The problem is when to switch the clock to accelerate the learning. The developments of robot learning can be triggered by not simply the internal clock but also the learning performance to what extent the robot improved its performance.

2.2 Robot Development Triggered by The Learning Performance

Here, we discuss a developmental approach triggered by the interaction between robot and environment. In this approach, the robot performance is improved by learning, and then the performance progress causes the development of the internal mechanisms.

It is known in physiology that the growth of visual cortex of cerebrum causes the development of visual capability [3, 7]. The sharpness of the retina images depends on the computational power of the visual cortex of cerebrum that processes the images. In other words, during the computational power of the visual cortex is at the lower level, the retina receives blurred images because the lens generates such images that the visual cortex can process within its computational power. Then, the images are carried to the visual cortex through the optic nerve, and the visual cortex learns a behavior through the interactions with environment. However, when the learning in the visual cortex advances and it has the higher computational power, the retina images would have fine resolution. The visual cortex learns a behavior using these fine images, as a result, the learning is accelerated and the final computational power is augmented.

It is supposed that these development caused by the interactions between a learner and an environment could be applicable to others, such as auditory and motional capabilities.

2.3 Environmental Development Triggered by The Given Clock

An environment including a caregiver may have developmental mechanisms as well as a robot. This kind of developmental approach is a method that accelerates the robot learning by changing the complexities of the environmental structure according to the pre-specified time schedule.

Education systems and social rules by laws in human society are examples of this kind of development. The educational establishments are scheduled by the pre-specified time clock, i.e. the academic year. In other words, the knowledges that children learn in the educational establishments changes from easy to difficult levels according to the clock. This shift helps the children to learn the knowledges step by step.

2.4 Environmental Development Triggered by The Learning Performance

Similar to the robot development approach caused by the learning process, there might be one more environmental development approach caused by the learning process. This approach is based on the knowledge from cognitive science, that is learning of children is accelerated by receiving appropriate problems or rewards from its environment [12]. These environmental developments are triggered by interactions between a learner and an environment. The environment observes the learner's performance through the learning process and changes the complexities of the environmental structure in accordance with the learner's performance.

Uchibe *et al.* [16] proposed a method to control the environmental complexity in accordance with the robot learning performance. In their method, the robot was provided the most difficult situation and estimated the full sets of state vector at first, where the environmental complexity was defined as the dimensions of the state vector. Their experimental results showed that the increasing of dimensions of the state vector according to the robot performance made the learning more efficient.

3 Learning of Joint Attention

Four kinds of relationships between a developmental agent and a trigger were explained above. These developmental patterns are generated separately or together. In this section, both robot and environmental development triggered by the learning performance are combined to apply to a learning problem of joint attention.

3.1 Joint Attention

Visual joint attention is defined as that an agent attends to the same object which another attends to. **Figure 2** shows the process of joint attention. In this situation, the robot is an agent of action. First, the robot observes the caregiver and estimates the direction of the caregiver's attention. Next, the robot turns the camera to the estimated direction and identifies the object that the caregiver attends to. These two steps are required for joint attention.

An ability of joint attention makes it possible to acquire many other social abilities [12]. For example, a robot with this ability can give and receive information with a caregiver, and it leads to acquisition of language and conceptions. Further, if it estimates the

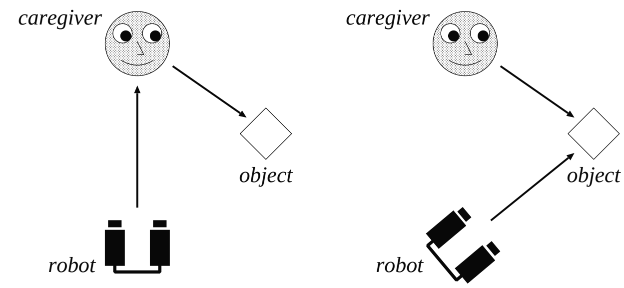


Figure 2: Joint attention between robot and caregiver

caregiver's behavioral intention from the direction of its attention, a robot could realize more smooth social communication.

Some researchers have attempted to realize a social communication between a human and a robot by constructing a joint attention mechanism inside the robot. Imai *et al.* [6] prepared a joint attention mechanism for their robot and evaluated its validity through a psychological experiment. In their case, a human was an agent of action. Their experimental result indicated that an eye-contact and an attention expression were significant factors in joint attention. Kozima [9, 10] discussed what is required for social communication between a human and a robot. He pointed out the necessity of a joint attention ability and an imitation one and constructed a simple joint attention mechanism inside their robot. Scassellati [14, 15] also built a joint attention mechanism toward theory of mind[1].

However, the joint attention mechanisms in these approaches were constructed by non-developmental ones, and the psychological implication between joint attention and development has not been involved.

3.2 Developmental Learning Mechanism for Joint Attention

We propose a learning mechanism for joint attention by a developmental approach shown in **Figure 3**. This has two multi-layered neural networks: one is for a learning robot and another for a caregiver. Each of them has a developmental mechanism triggered by interactions between a robot and a caregiver.

[1]Theory of mind is an ability to attribute beliefs, goals, and percept to other people [2].

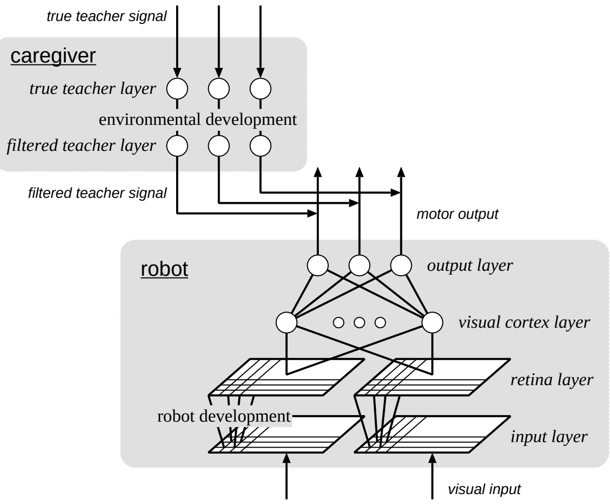


Figure 3: Developmental learning mechanism for joint attention

3.2.1 Neural Network for Robot

The neural network for a learning robot has four layers: an input layer, a retina layer, a visual cortex layer, and an output layer. The inputs for the network are left and right camera images when the robot observes the caregiver (the situation shown in Figure 2 (1)), and the outputs are motor commands to attend to the same object that the caregiver attends to (the situation shown in Figure 2 (2)).

The development of the visual mechanism inside the robot is implemented in the connection weights between the input layer and the retina one. **Figure 4** shows the appearances of robot and environmental development (environmental one is described in the next section). The connection weight W_i between the input and the retina layers is represented as a Gaussian spatial filter

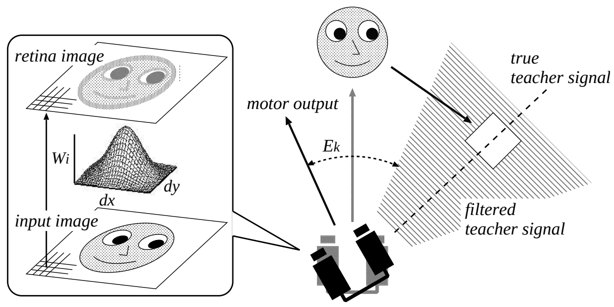
$$W_i = \exp\left(-\frac{(x-x_c)^2+(y-y_c)^2}{2{\sigma_k}^2}\right),$$

where x and y represent a position in an input image plane, x_c and y_c represent a position of a target pixel in that image, and σ_k is standard deviation at learning time step k that defines the sharpness of the spatial filter. When $E_{k-\Delta k} - E_k < \Delta E$, σ_k is updated by the following equation:

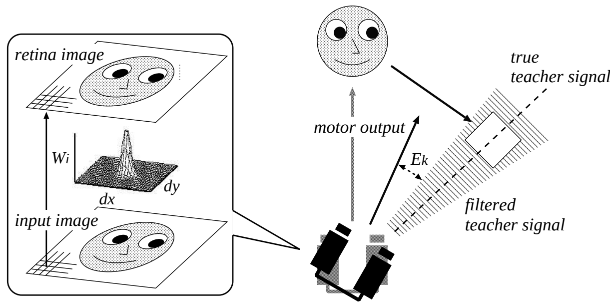
$$\sigma_k = \sigma_0\left(\frac{E_k - E_t}{E_0 - E_t}\right),$$

where E_k is the error between a motor output and a filtered teacher signal provided by the caregiver, E_0 is the initial error, and E_t is the tolerance after the learning. The change of σ_k can be regarded as visual development.

It is summarized as follows that the visual development is caused by the learning performance of the robot. The learning of other connection weights are conducted by the back propagation method.



(a) the early stage of learning



(b) the later stage of learning

Figure 4: Appearances of robot and environmental development

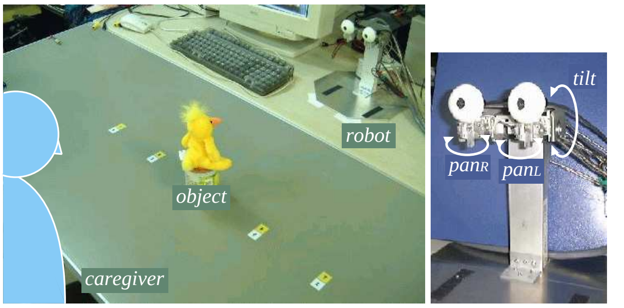
3.2.2 Neural Network for Caregiver

The neural network for a caregiver has two layers: a true teacher layer and a filtered teacher one. The caregiver provides teacher signals for the motor outputs of the robot and helps the robot learn.

The environmental development is implemented in the connection weights between two layers, and the connection works as a filter to control the resolution of the teacher signals. The appearance of this development is shown in Figure 4. The caregiver detects the true teacher signals and changes its resolution with the filter according to the learning advance of the robot, i.e. the error E_k. Then, the caregiver provides the filtered teacher signals to the robot. The filtered teacher signals are coarse value during the early stage of robot learning but become finer according to the learning process.

4 Experiment

4.1 Experimental Setup



(a) experimental environment (b) robot motion mechanism

Figure 5: An experimental setup for joint attention

The experimental environment is shown in **Figure 5** (a), and the motion mechanism of the robot is shown in (b). The robot is across the table, where an object is put on, from a caregiver. The positions of the robot and the caregiver are fixed. First, the caregiver attends to the object, and the robot observes the scene through its cameras. Note that it is assumed that the direction of the caregiver's attention is defined as its gaze direction, so that the inputs to the robot are images of only the caregiver's eye area. Then, the robots estimates the direction of caregiver's attention from the input images and turns its cameras to the estimated direction. The motor outputs are two independent pan angles for left and right cameras and one tilt angle. As a result of these process, the robot identifies the object, and joint attention is realized.

The learning of the robot's neural network is conducted by the back propagation method using the teacher signals that are obtained in various situation.

4.2 Experimental Results

The change of the average of normalized error per one output neuron of the robot's neural network is shown in **Figure 6**. The graph includes the result of the proposed developmental learning mechanism and the comparison result in which the caregiver does not help the robot to learn. The latter method is not that the caregiver provides the teacher signals to the robot,

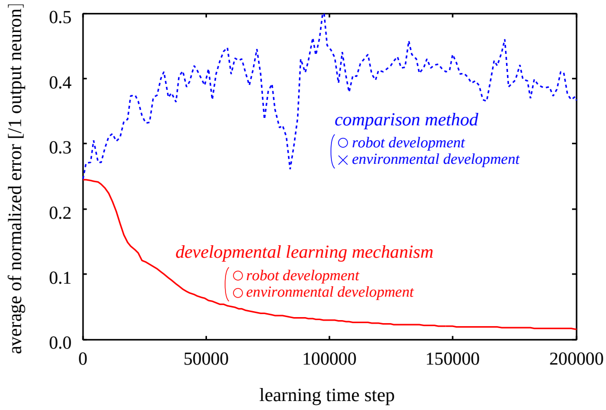


Figure 6: Error transition in learning process

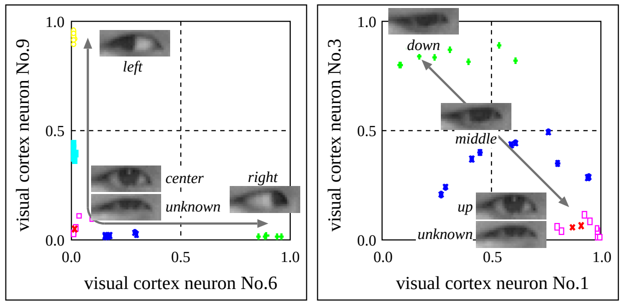


Figure 7: Relationships between input image and visual cortex neuron

but that the robot itself obtains the teacher signals. In other word, the robot obtains a teacher signal only when it is able to detect the object with its cameras as a result of its camera action.

The comparison between two methods shows that the proposed mechanism accelerates the learning of joint attention owing to the environmental development. In addition, it is shown that the learned mechanism has obtained an effective capability because it realizes joint attention wherever the object is put on.

It is analyzed how the visual cortex neurons in the learned neural network of robot extract features from the input images. **Figure 7** shows the relationships between the input images and the values of four visual cortex neurons out of ten that represent their features saliently. These relationships show that the neurons of No.6 and No.9 recognize the left-and-right direction of the caregiver's gaze with its iris position, and No.1 and No.3 recognize the up-and-down direction with its upper lid position. In addition, it is shown that these neurons recognize accurately the unknown and difficult input images when the caregiver attends to front side and squints its eyes. From these results, we may conclude that the visual cortex neurons in the learned neural network of robot appropriately extract the features of the input images.

5 Discussion toward Shared Attention

In this paper, the developmental approaches for a learning robot were categorized according to the relationships between what does develop and what does make the development. The developmental learning mechanism for joint attention that has two kinds of development: one is robot development triggered by the learning performance and another is environmental development triggered by that, was proposed. The experimental results showed that the proposed mechanism accelerated the learning of joint attention owing to the environmental development.

In the current method, the usefulness of the development seems limited, and it is necessary to investigate interactions between developmental patterns. As a future work, we are gain to attack shared attention problem in which the robot and the caregiver confirm the state of joint attention each other. Its capability enables the robot to behave more socially, which need a more developmental approach.

References

[1] Minoru Asada, Karl F. MacDorman, Hiroshi Ishiguro, and Yasuo Kuniyoshi. "Cognitive Developmental Robotics As a New Paradigm for the Design of Humanoid Robots," *Robotics and Autonomous System*, Vol. 37, pp. 185-193, 2001.

[2] Simon Baron-Cohen. *Mindblindness*, MIT Press, 1995.

[3] J. Gavin Bremner. *Infancy*, Blackwell, 1994.

[4] Melissa Dominguez and Robert A. Jacobs. "Visual Development and the Acquisition of Binocular Disparity Sensitivities," In *Proceedings of the Eighteenth International Conference on Machine Learning*, 2001.

[5] Jeffrey L. Elman, Elizabeth A. Bates, Mark H. Johnson, Annette Karmiloff-Smith, Domenico Parisi, and Kim Plunkett. *Rethinking Innateness: A connectionist perspective on development*, MIT Press, 1996.

[6] Michita Imai, Tetsuo Ono, and Hiroshi Ishiguro. "Physical Relation and Expression: Joint Attention for Human-Robot Interaction," In *Proceedings of 10th IEEE International Workshop on Robot and Human Communication*, 2001.

[7] Masami Iriki and Keisuke Toyama, editors. *Physiology 1 (in Japanese)*, chapter II-B, Bunkoudou, 1986.

[8] Mark H. Johnson. *Developmental Cognitive Neuroscience*, Blackwell, 1997.

[9] Hideki Kozima. "Attention-Sharing and Behavior-Sharing in Human-Robot Communication," In *Proceedings of the IEEE International Workshop on Robot and Human Communication*, pp. 9-14, 1998.

[10] Hideki Kozima and Hiroyuki Yano. "A Robot that Learns to Communicate with Human Caregivers," In *Proceedings of the First International Workshop on Epigenetic Robotics*, 2001.

[11] Giorgio Metta, Giulio Sandini, Lorenzo Natale, and Francesco Panerai. "Development and Robotics," In *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*, pp. 33-42, 2001.

[12] Chris Moore and Philip J. Dunham, editors. *Joint Attention: Its Origins and Role in Development*, Lawrence Erlbaum Associates, 1995.

[13] Elissa L. Newport. "Maturational Constraints on Language Learning," *Cognitive Science*, Vol. 14, pp. 11-28, 1990.

[14] Brian Scassellati. "Imitation and Mechanisms of Shared Attention: A Developmental Structure for Building Social Skills," In *Proceedings of the Autonomous Agents Workshop "Agents in Interaction - Acquiring Competence through Imitation"*, 1998.

[15] Brian Scassellati. "Theory of Mind for a Humanoid Robot," In *Proceedings of the First IEEE-RAS International Conference on Humanoid Robots*, 2000.

[16] Eiji Uchibe, Minoru Asada, and Koh Hosoda. "Environmental Complexity Control for Vision-Based Learning Mobile Robot," In *Proceedings of IEEE International Conference on Robotics and Automation*, pp. 1865-1870, 1998.