6

Emergence of Intelligence Through Reinforcement Learning with a Neural Network

Katsunari Shibata
Oita University
Japan

1. Introduction

"There exist many robots who faithfully execute given programs describing the way of image recognition, action planning, control and so forth. Can we call them intelligent robots?" In this chapter, the author who has had the above skepticism describes **the possibility of the emergence of intelligence or higher functions by the combination of Reinforcement Learning (RL) and a Neural Network (NN)**, reviewing his works up to now.

2. What is necessary in emergence of intelligence(1)(2)

If one student solves a very difficult problem without any difficulties facing a blackboard in a classroom, he/she looks very intelligent. However, if the student wrote the solution just as his/her mother had directed to him/her, the student cannot answer questions about the solution process, and cannot solve even similar or easier problems. Further interaction shows up less flexibility in his/her knowledge. When we see a humanoid robot is walking fast and smoothly, or when we see a communication robot responds appropriately to our talking, the robot looks an existence with intelligence. However, until now, the author has never met a robot who looks intelligent even after a long interaction with it. Why can't we provide enough knowledge for a robot to be really intelligent like humans?

When we compare the processing system between humans and robots, a big difference can be noticed easily. Our brain is massively parallel and cohesively flexible, while the robot process is usually modularized, sequential and not so flexible as shown in Fig. 1. As mentioned later, the massive parallelism and cohesive flexibility seem the origin of our very flexible behaviors considering many factors simultaneously without suffering from the "Frame Problem"(3)(4). The keys to figuring out the cause of the big difference are "modularization" and "consciousness", the author thinks.

When we survey the brain research and robotics research, we notice that the common fundamental strategy "functional modularization" lies in both. In the brain research, identification of the role of each area or region seems to be its destination. While, in the robotics research, the process is divided into some functional modules, such as recognition and control, and by sophisticating each functional module, high-functionality is realized in total.

At present, each developed function does not seem so flexible. Furthermore, as for the higher functions, unlike recognition that is located close to sensors, and unlike control that is located close to actuators, they are not located close to sensors or actuators. Therefore, either "what

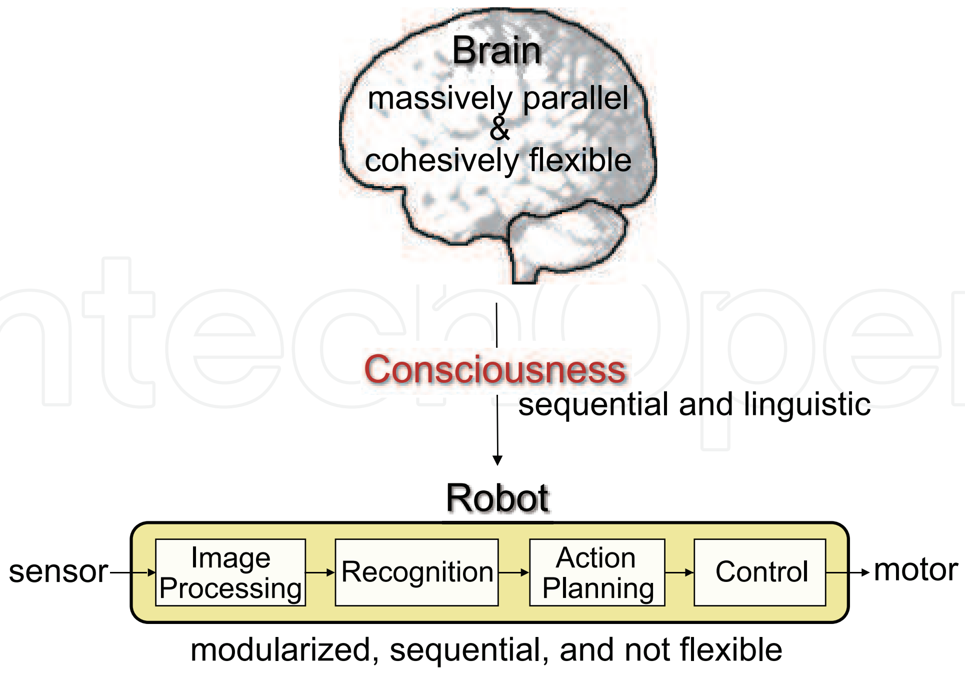


Fig. 1. Comparison of processing between humans and robots. The robot process is developed based on the understanding of the brain function through consciousness.

are the inputs" or "what are the outputs" is not predefined, and the both have to be designed by humans. However, since in higher functions such as language acquisition and formation of novel concept, very flexible function acquisition is required, it is difficult to decide even what are the inputs or what are the outputs in advance. Then the flexibility is deeply impaired when some definition of inputs and outputs are given. Accordingly, it is very difficult to develop a flexible higher function separately. The fundamental problem seems to lie in the isolation of each function from the system according to the "functional modularization" approach.

Why do we manage to modularize the entire process? It seems natural that researchers think the brain is too big to understand or develop. The fact that the brain seems to be divided into some areas by some major sulci probably promotes the modularization especially in the brain research. However, the author focuses on another more fundamental reason. That is the gap between the "brain" and "consciousness". It is said that the brain is a massively parallel system consisted of tens of billions or a hundred billion of neurons. We can see many reports showing that the flexibility of the brain is beyond expectation(5). Among them, a recent report is very impressive that the neuronal circuit remodeling for recovery after stroke spreads to the contralateral cortical hemisphere(6). On the other hand, our "consciousness" that is generated in the "brain" is sequential and its representation seems linguistic. So, it is difficult to represent and understand the function of the brain exactly as a massively parallel and very flexible system. Then by seeing the brain or brain functions through the frame of functional module, we reduce the amount of information, and try to understand it roughly by representing it linguistically. Thus, in the robot process that is designed in our conscious world, the functional modules are usually arranged in series as shown in Fig. 1.

Accordingly, it is thought to be impossible to understand the brain completely as long as through consciousness. However, since we do not notice the subconscious massively parallel processing directly, but notice only what we can see through consciousness, we cannot help but consider what we understand through the consciousness is all the processes that the brain is doing. We assume that since the brain is “doing”, the brain must be able to “understand” what the brain is doing. Then we expect that the brain will be understood by understanding each module individually, and also expect that human-like robots will be developed by building-block of sophisticated functional modules.

The existence of subconscious processes is undisputed even from the fact that the response of each “orientation selectivity cell” cannot be perceived. The phenomena such as “optical illusion” or “choice blindness”(7) can be considered as a result of the gap between “brain” and “consciousness”. When we walk up non-moving escalator stairs, although we understand that the escalator is not moving, we feel very strange as if we are pushed forward. This suggests the existence of subconscious compensation for the influence of escalator motion that occurs only when we are on an escalator. When we type a keyboard, many types of mistypings surprise us: neighbor key typing, character-order confusing, similar-pronunciation typing, similar-meaning word confusion and so forth. That suggests that our brain processing is more parallel than we think though we assume it difficult for our brain to consider many things in parallel because our consciousness is not parallel but sequential.

Even though imaging and electrical recording of brain activities might provide us sufficient information to understand the exact brain function, we would not be able to understand it by our sequential consciousness. The same output as our brain produces may be reproduced from the information, but complete reconstruction including flexibility is difficult to be realized, and without “understanding”, “transfer” to robots must be impossible.

From the above discussion, **in the research of intelligence or higher-functions, it is essential to notice that understanding the brain exactly or developing human-like robots based on the biased understanding is impossible. We have to drastically change the direction of research.** That is the first point of this chapter. Furthermore, to realize the comprehensive human-like process including the subconscious one, **it is required to introduce a massively parallel and very flexible learning system that can learn in harmony as a whole**. This is the second point.

Unlike the conventional sequential systems, Brooks advocated to introduce a parallel architecture and “Subsumption architecture” has been proposed(8). The agile and flexible motion produced by the architecture has made a certain role to avoid the “Frame Problem”(3)(4). However, he claimed the importance of understanding of complicated systems by the decomposition of them into parts. He suggests that functional modules called “layer” are arranged in parallel, and interfaces between layers are designed. However, as he mentioned by himself, the difficulties in interface design and scalability towards complicated systems are standing against us as a big wall.

Thinking again what a robot process should be, it should generate appropriate actuator outputs for achieving some purpose referring to its sensor signals; that is ”optimization” of the process from sensors to actuators under some criterion. If, as mentioned, the understanding through “consciousness” is limited actually, by prioritizing human understanding, it constrains robot’s functions and diminishes its flexibility unexpectedly.

For example, in action planning of robots or in explaining human arm movement(9), the term “desired trajectory” appears very often. The author thinks that the concept of “desired trajectory” emerges for human understanding. As the above example of the virtual force perceived on non-moving escalators, even for motion control, subconscious parallel and flexible processing must be performed in our human brain. In the case of human arm movement, commands for muscle fibers are the final output. So the entire process to move an arm is the process from sensors to muscle fibers. The inputs include not only the signals from muscle spindles, but also visual signals and so on, and the final commands should be produced by considering many factors in parallel. Our brain is so intelligent that the concept of “desired trajectory” is yielded to understand the motion control easily through “consciousness”, and the method of feedback control to achieve the desired trajectory has been developed. The author believes that the direct learning of the final actuator commands using a parallel learning system with many sensor inputs leads to acquisition of more flexible control from the viewpoint of the degrees of freedom. The author knows that the approach of giving a desired trajectory to each servo motor makes the design of biped robot walking easy, and actually, the author has not yet realized that a biped robot learns appropriate final motor commands for walking using non-servo motors. Nevertheless, the author conjectures that to realize flexible and agile motion control, a parallel learning system that learns appropriate final actuator commands is required. The feedback control does not need the desired trajectories, but needs the utilization of sensor signals to generate appropriate motions. That should be included in the parallel processing that is acquired or modified through learning.

Unless humans develop a robot’s process manually, “optimization” of the process should be put before “understandability” for humans. To generate better behaviors under given sensor signals, appropriate recognition from these sensor signals and memorization of necessary information also should be required. Accordingly, the optimization is not just “optimization of parameters”, but as a result, it has a possibility that a variety of functions emerge as necessary. If the optimization is the purpose of the system, avoiding the freedom and flexibility being spoiled by human interference, harmonious “optimization” of the whole system under a uniform criterion is preferable.

However, if the optimized system works only for the past experienced situations and does not work for future unknown situations, learning has no meaning. We will never receive exactly the same sensor signals as those we receive now. Nevertheless, in many cases, by making use of our past experiences, we can behave appropriately. This is our very superior ability, and it is essential to realize the ability to develop a robot with human-like intelligence. For that, the key issue is “abstraction” and “generalization” on the abstract space.

It is difficult to define the “abstraction” exactly, but it can be taken as extraction of important information and compression by cutting out unnecessary information. By ignoring trivial differences, the acquired knowledge becomes valid for other similar situations. Brooks has stated that “abstraction” is the essence of intelligence and the hard part of the problems being solved(8). For example, when we let a robot learn to hit back a tennis ball, first we may provide the position and size of the ball in the camera image to the robot. It is an essence of intelligence to discover that such information is important, but we usually provide it to the robot peremptorily. Suppose that to return the serve exactly as the robot intends, it is important to consider a subtle movement of the opponent’s racket in his/her serve motion. It is difficult to discover the fact through learning from a huge amount of sensor signals. However, if the images are preprocessed and only the ball position and size in the image are extracted and given, there is no way left for the robot to discover the importance of opponent movement by itself. As an alternative, if all pieces of information are given, it is inevitable to introduce a parallel and flexible system in which learning enables to extract meaningful information from huge amount of input signals. That has a possibility to solve the “Frame Problem” fundamentally, even though the problem remains; how to discover important information effectively.

A big problem in “abstraction” is how the criterion for “what is important information” is decided. “The degree of reproduction” of the original information from the compressed one can be considered easily. Concretely, the introduction of a sandglass(bottleneck)-type neural network(10)(11), and utilization of principal component analysis can be considered. A non-linear method(12), and the way of considering temporal relation(13)(14)(15) have been also proposed. However, it may not match the purpose of the system, and drastic reduction of information quantity through abstraction process cannot be expected because the huge amount of sensor signals has to be reproduced. Back to basics, it is desirable that the criterion for "abstraction" matches the criterion of "optimization" of the system. In other words, the way of "abstraction" should be acquired in the "optimization" of the system.

From the discussions, the third point in this chapter is **to put "optimization" of the system before the "understandability" for humans, and to optimize the whole of the massively parallel and cohesively flexible system under one uniform criterion**. Furthermore, **to eliminate too much interference by humans and to leave the intelligence development to the "optimization" by themselves** are included in the third point.

3. Marriage of reinforcement learning (RL) and neural network (NN)

The author has proposed the system that is consisted of one neural network (NN) from sensors to actuators, and is trained by the training signals generated based on reinforcement learning (RL). The NN is a parallel and flexible system. The NN requires training signals for learning, but if the training signals are given by humans, they become a constraint on the system optimization, and the NN cannot learn functions beyond what humans provides. RL can generate the training signals autonomously, and the NN can optimize the entire system flexibly and purposively based on the signals. Therefore, the system is optimized in total to generate appropriate motions for getting more reward and less punishment and also to evaluate states or actions appropriately. Sometimes it acquires unexpected abilities because of being free from the constraints that are unwillingly produced by its designer.

On the other hand, RL is usually taken as a learning for actions that appropriately generates the mapping from a sensor space to an action space. By introducing a NN, not only non-linear function approximation, but also acquisition of various function, including recognition and control, through learning to generate better actions are expected. When introducing a

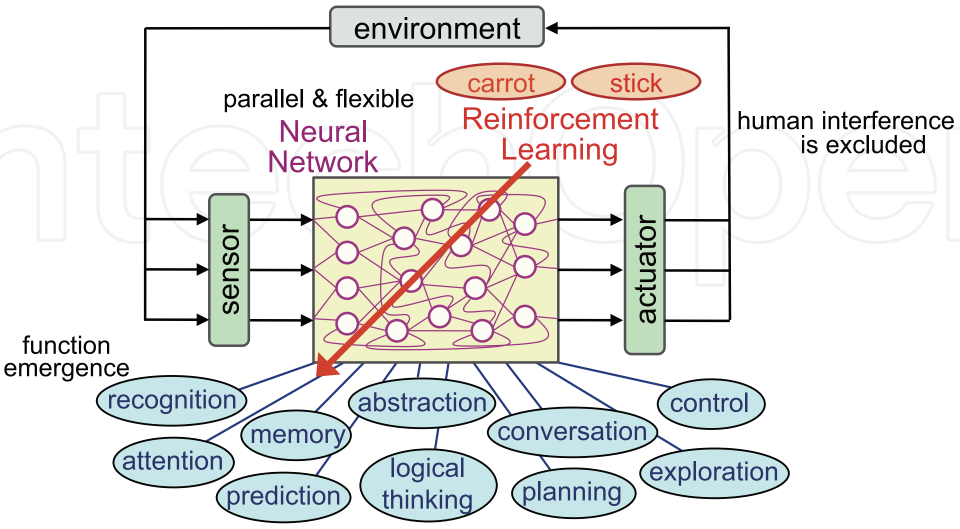


Fig. 2. The proposed learning system that is consisted of one neural network (NN) from sensors to actuators, and that is trained based on reinforcement learning (RL). Emergence of necessary functions based on the parallel and cohesively flexible learning is expected.

recurrent NN, functions that need memory or dynamics are expected to emerge. Thanking to its parallel and flexible learning system, flexible response to the real world considering various things in parallel is expected.

Comparing with the subsumption architecture that requires prior design of interactions among modules, flexibility, parallelism and harmony are stronger in the proposed system. The entire system changes flexibly, purposively and in harmony, and the interactions between neurons are formed autonomously. That is expected to solve the ”Frame Problem” fundamentally. It is also expected that in the hidden neurons, meaningful information is extracted among a huge amount of inputs, and the abstraction is consistent with the system optimization. Furthermore, from the viewpoint of higher functions, since the inside of the NN changes very flexibly, it is possible that necessary functions emerge without prior definition of “what higher function should be developed”, “what signals are inputs” or “what signals are outputs” of the higher function.

It is considered that “symbol emergence” and “logical thinking” is the most typical higher functions of humans. Symbol processing has been separately considered from pattern processing, linking to the difference of function between right and left hemispheres of the brain(16). NNs have been considered as a system for pattern processing. The idea has disturbed the investigation of symbol processing with a NN. Until now, it is not clear how these functions emerge, or which kind of necessity drives the emergence of these functions. However, if symbol processing emerges in an artificial NN with sensory signal inputs, it is expected that the clear boundary between symbols and patterns disappears, and the “Symbol Grounding Problem”(17) is solved. It might be no doubt that the human brain, which is consisted of a natural NN, realizes our logical thinking. Even though it is said that the function of the right hemisphere is different from that of the left one, one hemisphere looks very similar to the other.

At present, in RL, a NN is positioned mainly as a nonlinear function approximator to avoid the “curse of dimensionality” problem(18), and the expectation towards purposive and flexible function emergence based on parallel processing has not been seen. A NN was used in an inverted pendulum task(19) and Backgammon game(20), but since the instability in RL was pointed in 1995(21), function approximators with local representation units such as NGnet (Normalized Gaussian network)(22) are used in many cases(23). In the famous book as a sort of bible of RL(18), little space is devoted for the NN.

However, the very autonomous and flexible learning of the NN is surprising. Even though each neuron performs output computation and learning (weight modification) in a uniform way, the autonomous division of roles among hidden neurons through learning and purposive acquisition of necessary internal representation to realize required input-output relations make us feel the possibility of not only function approximation, but also function emergence and “intelligence”.

As mentioned, it is said that the combination of RL and a NN destabilizes the learning(21). In RBF network(24) including NGnet(22) or tile coding (CMAC)(25)(26), since a continuous space is divided softly into local states, learning is performed only in one of the local states and that makes learning stable. However, on the other hand, they have no way to represent more global states that integrate the local states. The sigmoid-based regular NN has an ability to reconstruct a useful state space by integrating input signals each of which represents local information, and through the generalization on the internal state space, the knowledge acquired in past experiences can be utilized in other situations(27)(28).

The author shows that when each input signal represents local information, learning becomes stable even using a regular sigmoid-type NN(29). Sensor signals such as visual signals originally represent local information, and so learning is stable when sensor signals are the inputs of the NN. On the contrary, when the sensor signals are put into a NN after converting them to a global representation, learning sometimes became unstable. If the input signals represent local information, a new state space that is good for computing the outputs is reconstructed in the hidden layer flexibly and purposively by integrating the input signals. Therefore, both learning stability and flexible acquisition of internal representation can be realized. For the case that input signals represent global information, Gauss-Sigmoid NN has been proposed where the input signals are passed to the sigmoid-type regular NN after localization by a Gaussian layer(30).

On the other hand, in the recent researches of learning-based intelligence, "prediction" of a future state from the past and present states and actions has been focused on because it can be learned autonomously by using the actual future state as the training signals (13)(31)(32)(14)(33). However, it is difficult and also seems meaningless to predict all of the huge amount of sensor signals. Then it becomes a big problem how to decide "what information at what timing should be predicted". This is similar to the previous discussion about "abstraction". A way to discover the prediction target from the aspect of linear independence has been proposed(34). However, same as the discussion about "abstraction", thinking about the purposive property and consistency with the system purpose, "prediction" should be considered in RL. The author's group shows that through learning a prediction-required task using a recurrent NN, the function of "prediction" emerges(35) as described later.

Next, how to learn a NN based on RL is described. In the case of actor-critic(36), one critic output unit and the same number of actor output units as the actuators are prepared. Actuators are operated actually according to the sum of the actor outputs $\mathbf{O}a(\mathbf{S}_t)$ for the sensor signal inputs $\mathbf{S}_t$ and random numbers $\mathbf{rnd}_t$ as trial and error factors. Then the training signals for critic output Tc_t and for actor outputs $\mathbf{T}a_t$ are computed using the reward r_{t+1} obtained by the motion and critic output $Oc(\mathbf{S}_{t+1})$ for the new sensor signals $\mathbf{S}_{t+1}$ as

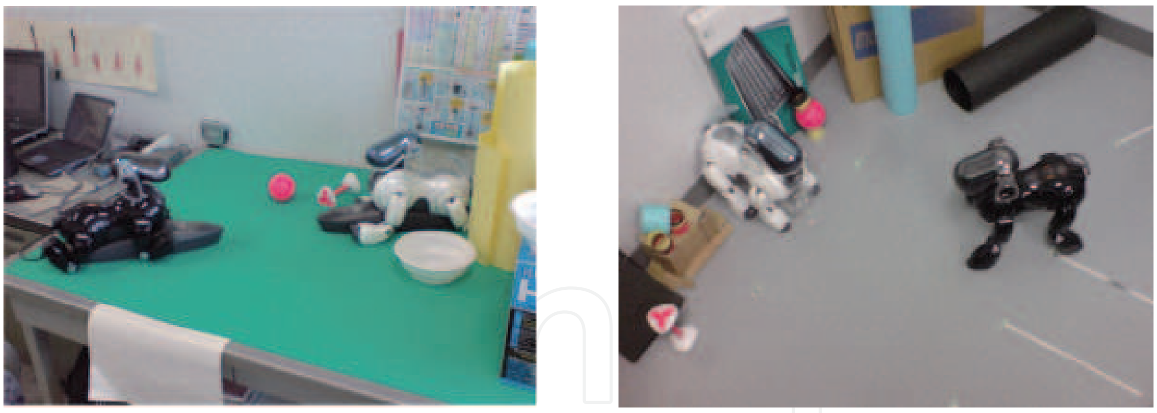
$$\begin{aligned} Tc_t &= Oc(\mathbf{S}_t) + \hat{r}_t = r_{t+1} + \gamma Oc(\mathbf{S}_{t+1}) & (1) \\ \mathbf{T}a_t &= \mathbf{O}a(\mathbf{S}_t) + \alpha \hat{r}_t \mathbf{rnd}_t & (2) \\ \hat{r}_t &= r_{t+1} + \gamma Oc(\mathbf{S}_{t+1}) - Oc(\mathbf{S}_t) & (3) \end{aligned}$$

where $\hat{r}$ indicates TD-error, γ indicates a discount factor and α indicates a constant. After that, the sensor signals $\mathbf{S}_t$ at the time t are provided as inputs again, and the NN is trained by the BP (Error Back Propagation) learning(10) using the training signals as above. If the neural network is recurrent-type, BPTT(Back Propagation Through Time)(10) can be used.

On the other hand, in the case of Q-learning(37), the same number of output units as the actions are prepared in the NN, and each output $O_a(\mathbf{S}_t)$ is used as the Q-value for the corresponding action a. Using the maximum Q-value for the new sensor signals $\mathbf{S}_{t+1}$ perceived after the selected action a_t, the training signal for the output for the action a_t at the time t as

$$T_{a_t,t} = r_{t+1} + \gamma(\max_a O_a(\mathbf{S}_{t+1})). \quad (4)$$

Then, after input of the sensor signals $\mathbf{S}_t$ and forward computation, only the output for the selected action a_t is trained. When the value range of critic, actor or Q-values is different from



(a) head rotation task

(b) walkling task

Fig. 3. Two learning tasks using two AIBO robots.

the value range of the NN output, linear transformation can be applied. The learning is very simple and general, and can be widely applied to various tasks.

4. Some examples of learning

In this section, some examples are introduced in each of which emergence of recognition in a real-world-like environment, memory, prediction, abstraction or communication is aimed. To purely see the abilities and limitations of function emergence, the author has intentionally insisted in “learning from scratch”. Therefore, the required functions are still very simple, but it is confirmed that various functions except for logical thinking emerge in a NN through RL almost from scratch. It must be the honest impression for the readers who read this chapter up here that only by connecting from sensors to actuators by a flat NN and training it very simply and generally only from a scalar reinforcement signal, it is just too much for a robot or agent to solve a difficult task or to acquire higher functions from scratch. The author are pleased if the readers feel a new tide that is different from the previous approach in robotics research, and the possibility of function emergence by the couple of RL and a NN. In order to see them from the viewpoint of function emergence, the readers are asked to focus on the ratio of acquired function versus prior knowledge and also the flexibility of the function. The feasibility of the proposed approach and method will be discussed in the next section. The details of the following examples can be referred to the reference for each.

4.1 Learning of flexible recognition in a real-world-like environment

We executed two experiments using two AIBO robots as shown in Fig.3. In one of them(2) that is named “head rotation task”, two AIBOs were put face-to-face as shown in Fig.3(a). The head can be one of 9 discrete states with the interval of 5 degree. The AIBO can take one of the three actions: “rotate right”, “rotate left” and “bark”. When it barks capturing the other AIBO at the center of the image, a reward is given, on the other hand, when it barks in the other 8 states, a penalty is given. In the second task(38) named “walking task”, the AIBO is put randomly at each trial, and walks as shown in Fig.3(b). When it kisses the other AIBO, a reward is given, and on the other hand, when it loses the other AIBO, a penalty is given. The action can be one of the three actions: “go forward”, “turn right”, and “turn left”. In this task, state space is continuous, and the orientation and size of the other AIBO in the image are varied.

As shown in Fig. 4 that is for the case of walking task, the 52 × 40 pixel color image that is captured by the camera mounted at the nose of the AIBO are the input of the NN in both tasks. A total of 6240 signals are given as inputs without any information about the pixel location. The NN has 5 layers, and the numbers of neurons are 6240-600-150-40-3 from the input layer to the output layer. The network is trained by the training signals generated based on Q-learning. The lighting condition and background are changed during learning. Figure 5 shows some sample images in the second task. Since no information about the task is given and no knowledge to recognize the AIBO is given, it is expected for the readers to understand that the learning is not so easy.

The success rate reached more than 90% in the first task after 20,000 episodes of learning, and around 80 or 90% in the second task after 4,000 episodes of learning with additional learning using the experienced episode. Of course, that is far inferior than when a human do the same task, but it is interesting that without giving any knowledge about the task or image recognition, image recognition of the AIBO can be acquired to some extent through the learning with only reward and punishment. What are even more interesting can be found in the analysis of the internal representation of the NN.

At first, the 6240 connection weights from the input neurons to each of the 600 lowest hidden neurons were observed as a color image with 52 × 40 pixels. When the weights are normalized to a value from 0 to 255, the image looks almost random because of the random initial weights. Then the weight change during learning is normalized to a value from 0 to 255. For example, when the connection weight from the red signal of a pixel increases through learning, the corresponding pixel looks redder, and when it decreases, the pixel looks less red. Figure 6 (a) shows some images each of which represents the change of the connection weights to one of the 600 lowest hidden neurons in the head rotation task. In the head rotation task, one or more AIBO's figures can be found in the image although there are two ways of the

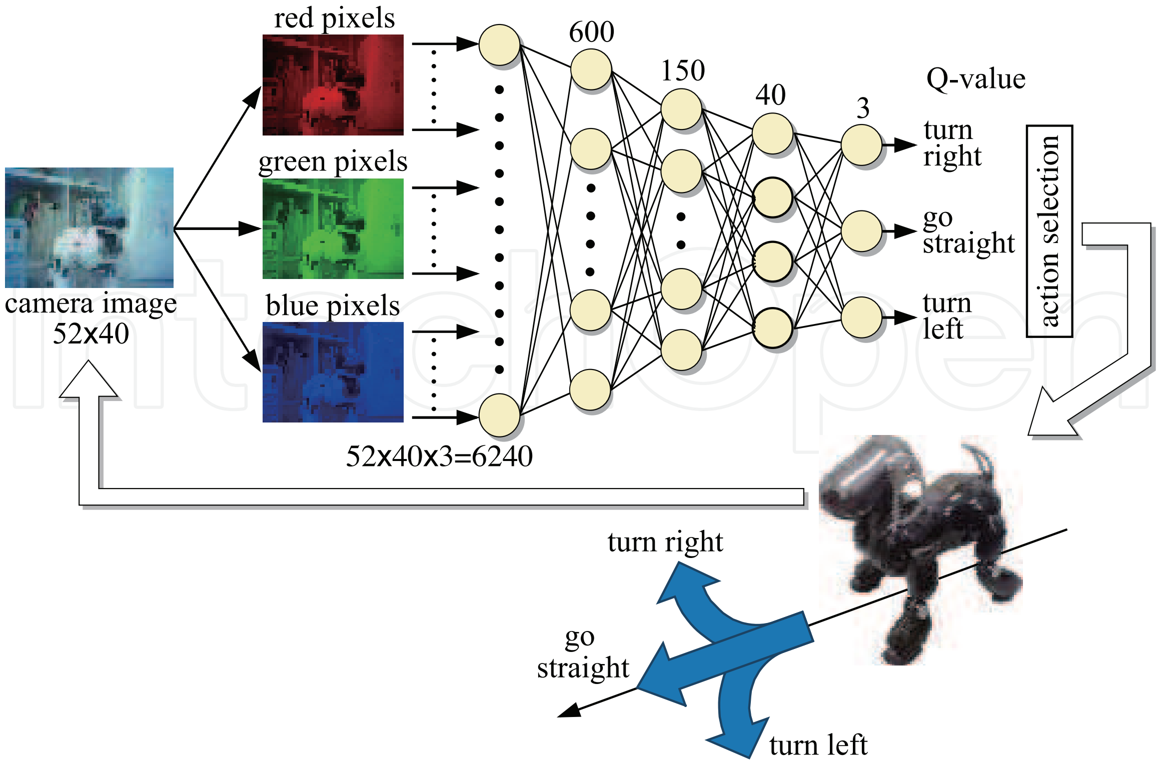


Fig. 4. The learning system and flow of the signals in the AIBO walking task.

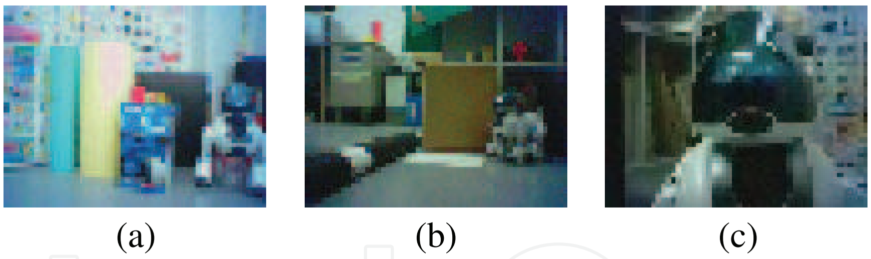


Fig. 5. Variety of images whose pixel values are directly put into a NN as input signals (2).

view of AIBO; positive one and negative one. Because the AIBO is located at only one of the 9 locations, it seems natural that AIBO figures can be found in the image, but the place where the AIBO figure appears is different among the hidden neurons. In (a-1), the neuron seems to detect the AIBO at the left of the image, and in (a-5), the neuron seems to contribute to make a contrast whether the AIBO is located at the center or not. It is interesting that autonomous division of roles among hidden neurons emerged just through RL. It is possible that the contrast by simultaneous existence of positive and negative figures contributes to eliminating the influence of lighting condition.

Figure 6 (b) shows the weight change from the view of a middle hidden neuron. The image is the average of the images of 600 lowest hidden neurons weighted by the connection weights from the lowest hidden neurons to the middle hidden neuron. In many of the images of middle hidden neurons, AIBO figure looks fatter. It is possible that that absorbs the inaccurate head control of AIBO due to the use of a real robot.

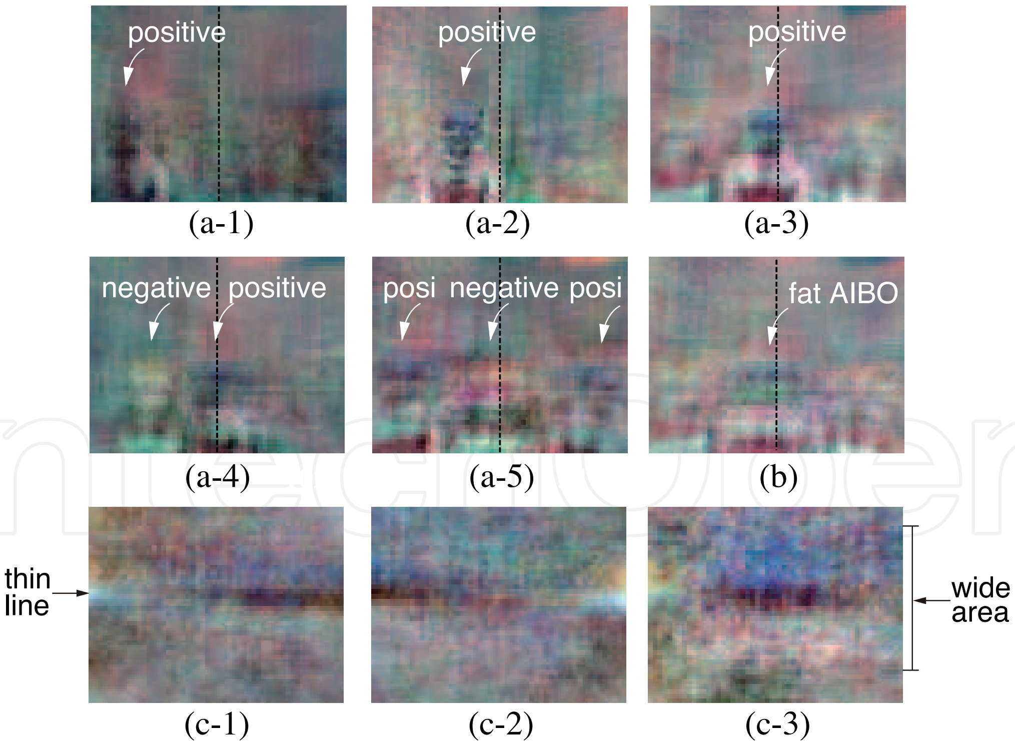
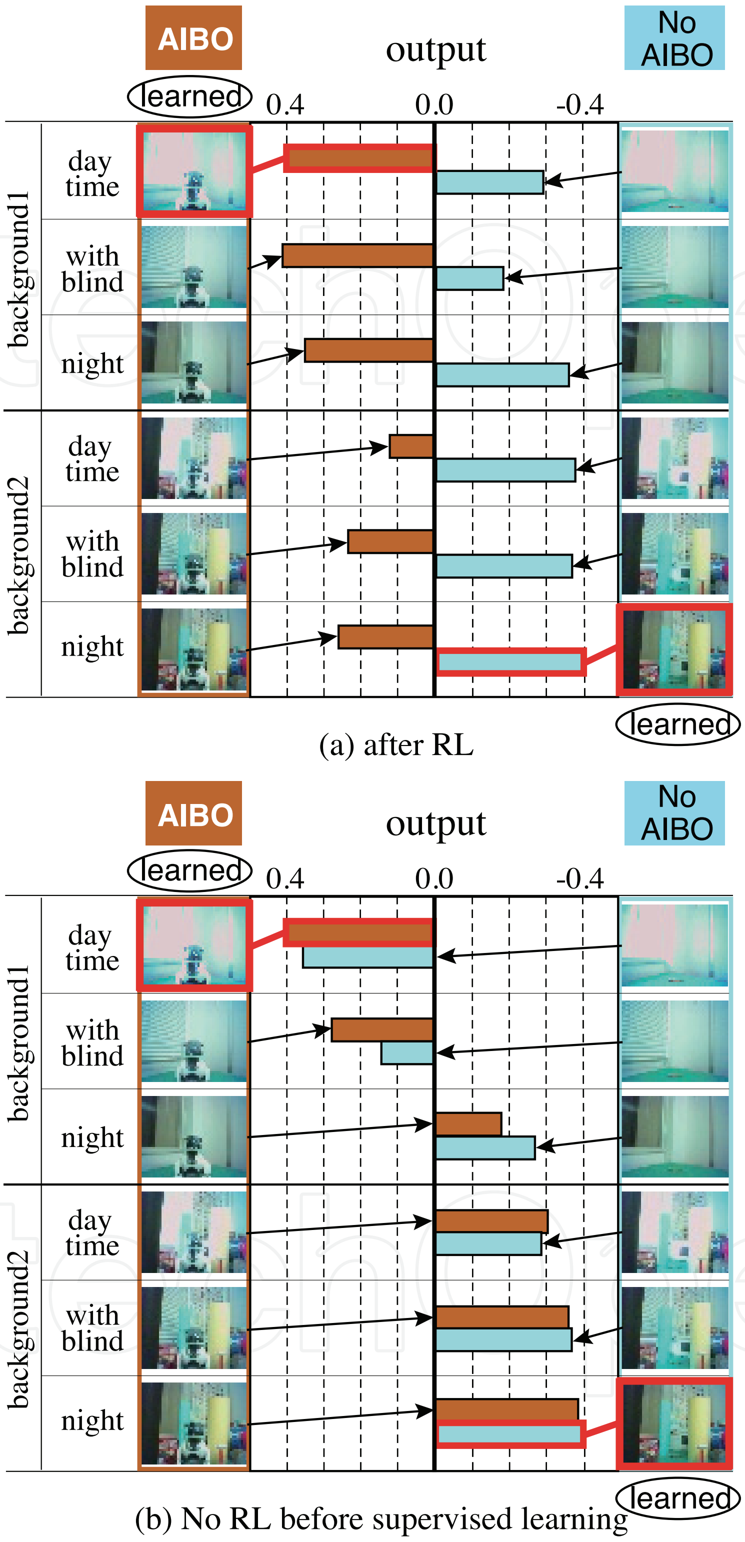


Fig. 6. Images to represent weight change in some hidden neurons during RL. (a) weight change in 5 lowest hidden neurons during the head rotation task, (b) weight change in a middle hidden neuron during the head rotation task, (c) weight change in 3 middle hidden neurons during the walking task.



(a) after RL

(b) No RL before supervised learning

Fig. 7. To see the change of internal representation through RL in the head rotation task, after supervised learning of 2 learning patterns, which are surrounded by a red frame in this figure, the output for 10 test patterns are compared between the NN after RL and that before RL. After RL, the output depends deeply on whether the AIBO exists or not, while before RL, the output depends on the brightness and background.

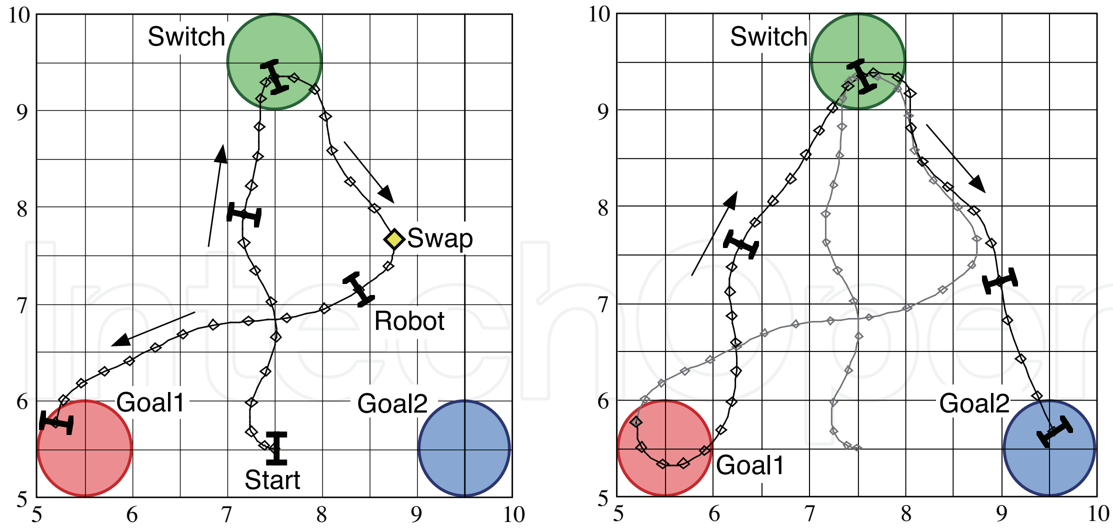
In the case of walking task, most of the weight images for the lowest hidden neurons are very vague, and in some of them, various vague AIBO figure can be found. Figure 6 (c) shows weight images for 3 middle hidden neurons. Unlike the case of head-rotation task, black and white thin line arranged one above the other can be seen in many images. In this task, since the location and orientation of the target AIBO are not limited, the walking AIBO seems to learn to recognize the target AIBO more effectively by focusing the black area of its face and the white area around its chin or of its body. The neurons as shown in Fig. (c-1)(c-2) seem to contribute to detecting lateral location of AIBO. While the neuron as shown in Fig. (c-3), which has a wide dark-blue area and wide white area, seems to contribute to detecting that the AIBO is closely located, because when the AIBO is close to the other AIBO, the face occupies a wide area of the camera image as shown in Figure 5(c).

It is interesting that the acquired way to recognize the AIBO is different between the two tasks. In the head rotation task, the recognition seems based on pattern matching, while in the walking task, it seems based on feature extraction.

One more analysis about the internal representation is reported. Here, the acquisition of internal representation of AIBO recognition not depending on the light condition or background is shown in the head rotation task. After RL, one output neuron is added with all the connection weights from the highest hidden neurons being 0.0. As shown in Fig. 7, 12 images are prepared. In the supervised learning phase, 2 images are presented alternately, and the network is trained by supervised learning with the training signal 0.4 for one image and -0.4 for the other. The output function of each hidden and output neuron is a sigmoid function whose value ranges from -0.5 to 0.5. One of the images is taken in daytime and bright, and the other is taken at night under fluorescent light and dark, and the background is also different. In the bright image, the AIBO exists in the center, and in the dark one, there is no AIBO. In 6 images, there is the AIBO at the center, and in the other 6 images there is no AIBO. Each of the 6 images with AIBO has a corresponding image in the other 6 images. The corresponding images are captured with the same lighting condition and background, and only the difference is whether the AIBO exists or not. In the 6 images, each 3 images have the same background, but different lighting condition. 3 lighting conditions are “daytime”, “daytime with blind” and “night”. The output of the NN is observed when each of 12 images is given as inputs. For comparison, the output is also observed when using the NN before RL. Figure 7 shows the output for the 12 images. After RL, the output changes mainly according to whether the AIBO exists or not, while before RL, the outputs is not influenced so much by the existence of the AIBO but is influenced by the lighting conditions and background. When the lighting condition or background is different between two images, the distance between them becomes larger than when the existence of AIBO is different. This result suggests that through RL, the NN acquired the internal representation of AIBO recognition not depending on the lighting conditions and backgrounds.

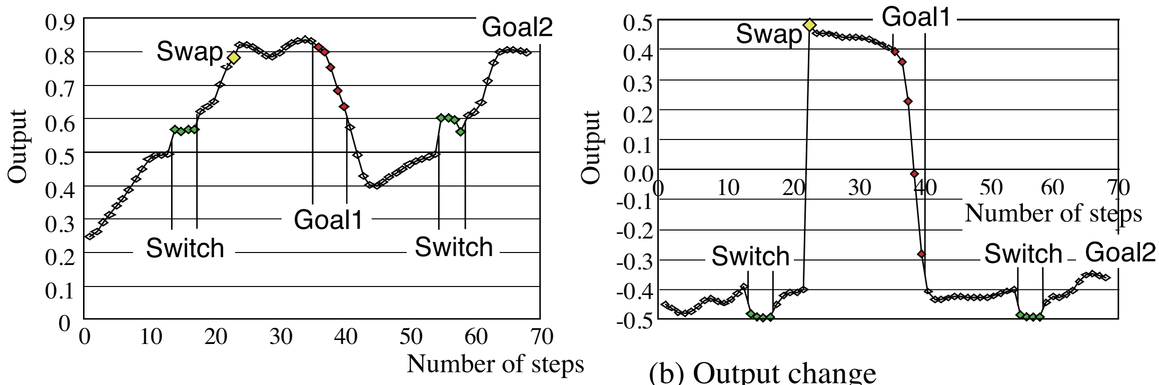
4.2 Learning of memory with a recurrent neural network (RNN)(39)

Next, learning of a memory-required task using a recurrent neural network (RNN) is reported. In this task, a wheel-type robot can get a reward when it goes to the correct one of two possible goals. One switch and two goals are located randomly, and the robot can perceive two flag signals only on the switch. When the flag1 is one, the correct goal is goal1, and on the other hand, when the flag2 is one, the correct goal is goal2. The inputs of the NN are the signals representing angle and distance to each of the switch and goals, distance to the wall, and also two flag signals. For the continuous motion, actor-critic is employed, and for the necessity of



(a) Former Behavior (b) Latter Behavior

Fig. 8. An interesting memory-based behavior acquired through RL. The robot gets a reward when it goes to the correct goal that can be known from the flag signals perceived only on the switch. In this episode, on the way to the goal2, the outputs of all the hidden neurons are swapped to the previously stored ones.



(a) Change of critic output

(b) Output change of a type 1 hidden neuron

Fig. 9. The change of the outputs of critic and a type 1 hidden neuron during the episode as shown in Fig. 8.

keeping the flag signals, a RNN is used. When it collides with the wall, when the robot comes to the goal without going to the switch, and also when it comes to the incorrect goal, a penalty is given to the robot.

After learning, the robot went to the switch at first, and then went to the correct goal that was known from the flag signals on the switch. When the output of each hidden neuron was observed, three types of hidden neurons that contribute to memory of necessary information could be found. Type1 neurons kept the flag1 signal, type2 neurons kept the flag2 signal, and type3 neurons kept either of the flag1 or flag2 signal is one. After the robot perceived that the flag1 was one on the switch, the output of one of the type1 neurons was reset to the initial value on the way to the goal1. Then the output soon returned to the value representing that the flag1 signal was one. That represents that a fixed-point attractor was formed through learning; in other words, associative memory function emerged through learning.

When some outputs of hidden neurons were manipulated, the robot showed interesting behaviors as shown in Fig. 8. The outputs of all the hidden neurons after perceiving flag1 being one were stored on the way to the goal1. At the next episode, the robot began to move from the same location with the same arrangement of switch and goals as the previous episode, but in this episode, the flag2 was on. After perceiving the flag signals on the switch, the robot approached the goal2. On the way to the goal2, the outputs of the hidden neurons were swapped by the ones stored on the way to the goal1 at the previous episode. Then the robot changed its traveling direction suddenly to the goal1. However, in this case, the goal1 was not the real goal, the robot could not get a reward and the episode did not terminate even though the robot reached the goal1. Surprisingly, the robot then went to the switch again, and finally went to the goal2 after perceiving the flag signals again. The NN during the behavior was investigated. As shown in Fig. 9 (a), the critic output decreased suddenly when the robot arrived at the goal1 as if the robot understood the goal1 is not the real goal. As shown in Fig. 9(b), a type 1 neuron kept a high value after the value swapping, but when it reaches the goal1, the value decreased suddenly. It is interesting to remind us the person who returns to check something again when they get worried.

4.3 Learning of prediction(35)

The next example shows the emergence of prediction and memory of continuous information through RL. In this task, as shown in Fig. 10, an object starts from the left end ($x = 0$) of the area, and its velocity and angle to go are decided randomly at each episode. The object can be seen until it reaches $x = 3$, but it often becomes invisible at $x > 3$ or a part of $x > 3$. The velocity of the object is decreased when it reflects at a wall. The agent moves along the line of $x = 6$, and decides the timing to catch the object. As input, the agent can receive the signals representing the object or the agent location locally. When the object cannot be seen, the signals for the object location are all 0.0. The agent can choose one of four possible actions; those are "go up", "go down", "stay" and "catch the object". If the agent can catch the object at the place close to the object, the agent can get a reward. The reward is larger when the object is closer to the agent. When it selects catch action away from the object, or does not select catch action before the object reaches the right end of the area, a small penalty is imposed to the agent. In this case, a RNN and Q-learning are used.

After learning, the agent became to catch the object appropriately even though the average reward was a little bit less than the ideal value. It is thought that the complete mechanism of prediction and memory cannot be understood easily, but a possible mechanism was found.

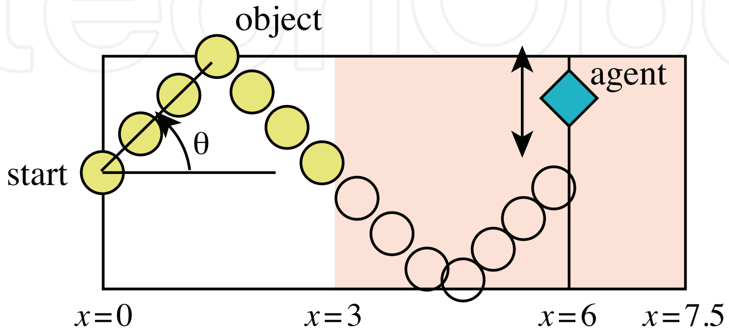


Fig. 10. Prediction task. The object velocity and traveling direction are randomly chosen at each episode, and the object becomes invisible in the range of $x > 3$ or a part of the range. The agent has to predict the object motion to catch it at an appropriate place and timing.

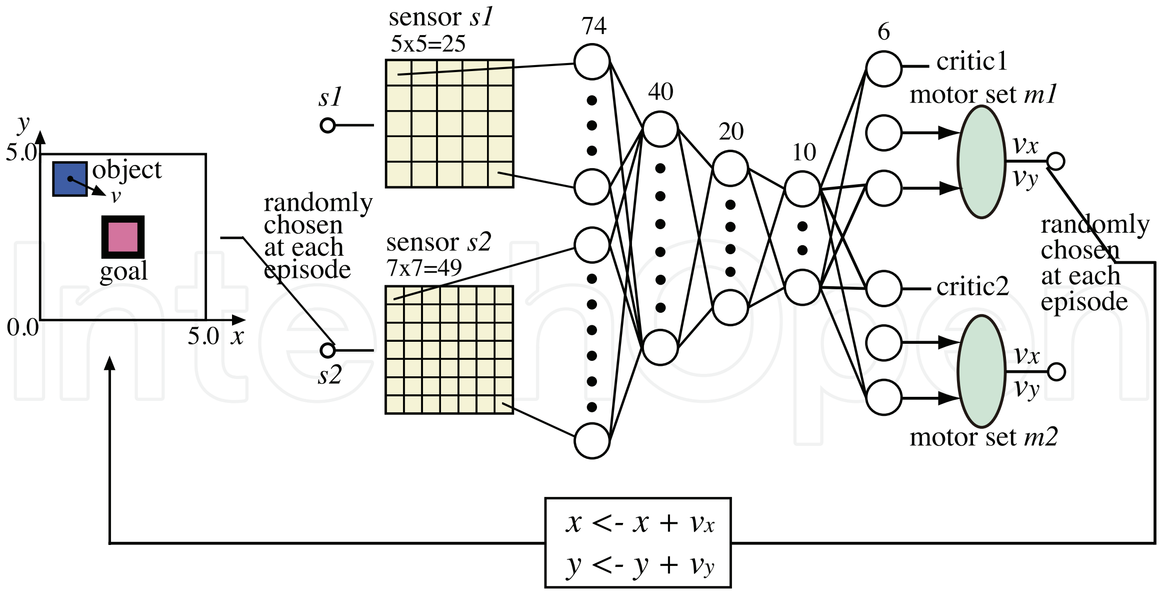


Fig. 11. Abstraction task. One of the two sensors and one of the two motor sets are randomly chosen at each episode. Learning is done for 3 of 4 combinations and the performance of s2-m2 combination is observed.

The agent detected the velocity in the x-direction from the period of staying specific area just before $x = 3$. The predicted value that should be memorized until the agent catches the object is not binary. If the memorized information is binary, the RNN can learn to memorize it by forming a bistable dynamics. However, in this case, it was observed that hidden neurons represent non-binary information for a while, and then relay it to the other neurons until the memorized information is utilized. Similar memory function also emerged in the learning of deterministic exploration task(40).

4.4 Learning of abstraction and knowledge transfer(41)

As mentioned before, one of the human's superior functions is abstraction and knowledge transfer. In our life, completely the same situation or sensor signals at present will never appear in the future. Nevertheless, we can behave appropriately in many cases utilizing our past experiences.

Then a simple task as shown in Fig. 11 is introduced to show the purposive abstraction clearly. An agent has two sensors s1 and s2, and two motor sets m1 and m2. Either sensor can perceive the positional relation of the object from the goal though the way of perception is different between the two sensors. In the same way, either motor set can control the object location though the way of control is different between the two motor sets. At each episode, one sensor and one motor set are selected randomly. The sensor signals from the non-selected sensor are all 0.0. From the view of NN, even though the positional relation of the object and goal is completely the same, the sensor signals from the two sensors are completely different. When the object reaches the goal area, the agent can get a reward.

Three of four sensor-motor combinations are used alternately in learning. The other combination s2-m2 is not used in learning, but just the performance is observed. Figure 12 shows the learning curve. The performance of s2-m2 combination is getting better even though no learning is done for the combination. When two of the four combinations were used in learning, the performance of either of the non-trained combinations did not get better. Through the learning for s1-m1 and s2-m1 combinations, the common representation not

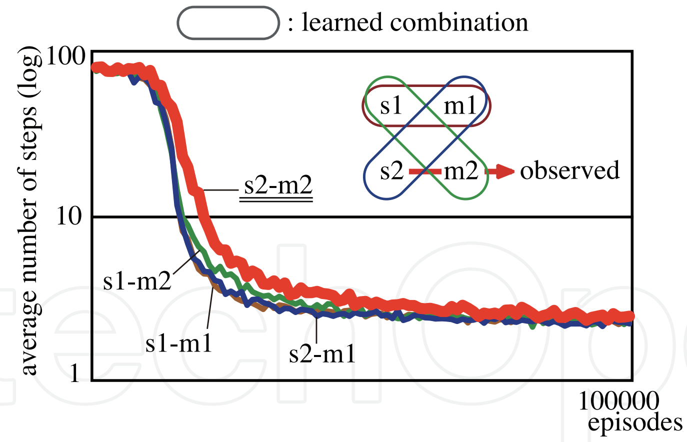


Fig. 12. Learning curve for 4 sensor-motor combinations. The performance is getting better for the s2-m2 combination for which no learning is done.

depending on the used sensor can be obtained in the hidden layer because when the desired output is similar, the hidden representation becomes similar after learning even though input signals are different(27)(28). Furthermore, through the learning for s1-m2 combination, the mapping from the common representation to the motor set m2 can be acquired. Since the conventional abstraction methods(10)-(15) abstract input signals by watching only the input signals or the time series of them as mentioned, this purposive abstraction obtained by the proposed method cannot be acquired by the conventional ones.

4.5 Learning of communication and binarization of signals(42)

Last report is concerning about one of the typical higher functions: communication. As shown in Fig. 13, one of two agents can transmit two continuous-valued signals sequentially based on the perceived location of the other agent, but cannot move by itself. The other agent can receive the transmitted signals and move, but it cannot see the other agent. When they meet, that means the receiver agent reaches close to the transmitter agent, they can both get a reward. They are in one-dimensional space, and the receiver agent can reach the transmitter in one step from anywhere when it takes an appropriate motion. As shown in Fig. 14, each of them has a RNN, and the network works to generate two sequential signals or to generate appropriate motion from the two sequential signals received. Furthermore, some noise is added to the communication signals. If the noise level is large, the information of the original signals is lost when the signal has an analog representation.

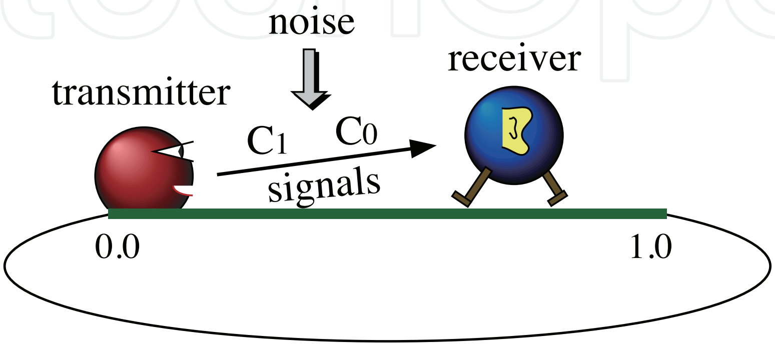


Fig. 13. Communication learning task. Two signals are transmitted sequentially, and the receiver moves according to the signals. Random noise is added to the signals.

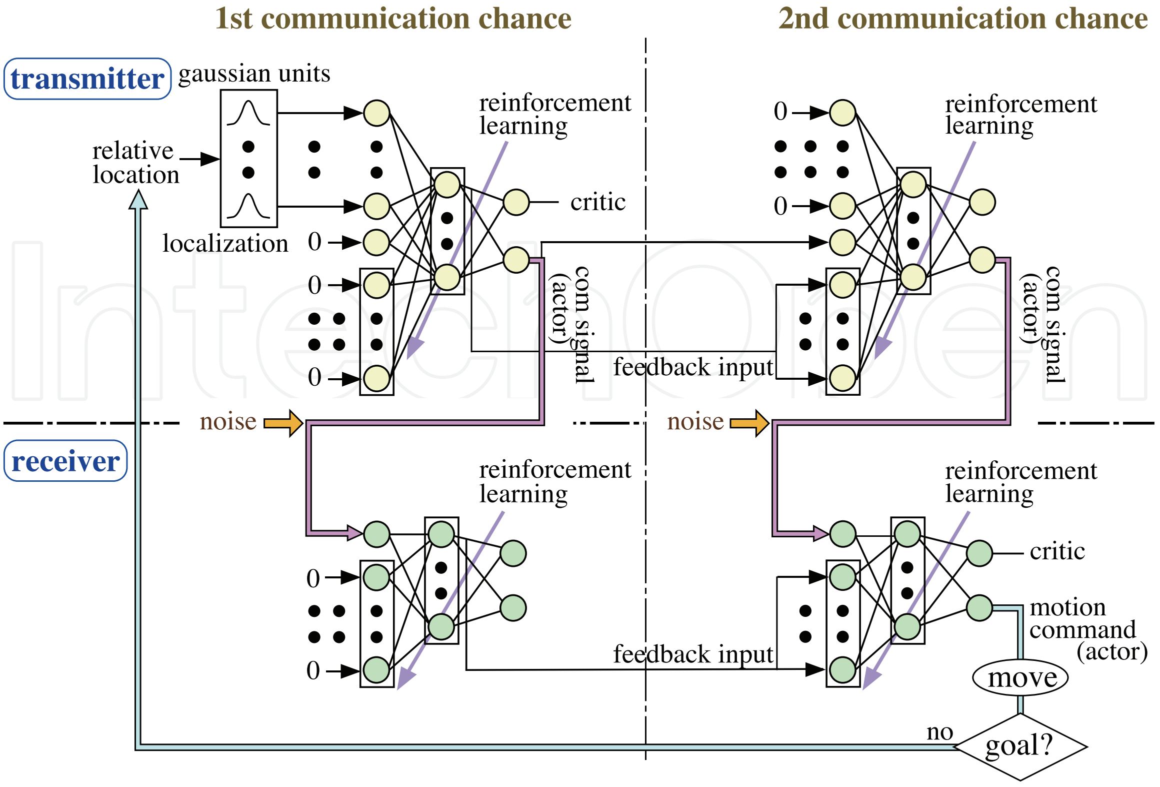


Fig. 14. The architecture and signal flow in the communication learning. Each agent has a recurrent neural network.

After learning, they could meet in one step in almost all cases. Figure 15 shows two communication signals generated by the transmitter as a function of the receiver's location in two cases when no noise is added and when a large level of noise is added. It is interesting that when the agents learned in the noisy environment, the signals get to have an almost binary representation even though no one told them to binarize the signals. If the representation is binary, the original signal before adding the noise can be restored. In this task, division of the receiver's location into 4 states is enough to generate a motion to reach the goal in

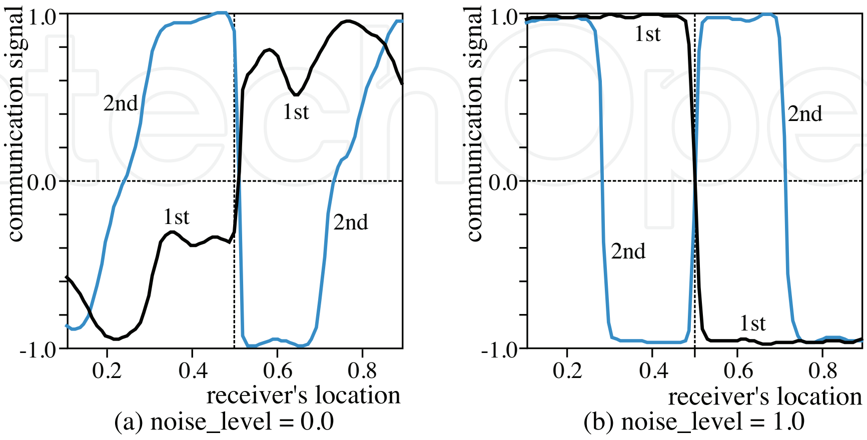


Fig. 15. Communication signals acquired through learning. In the case of large noise level(b), the signals have almost binary representation. That means that 2bit serial communication was established through RL.

one step. The sequential two signals represent different information, and the 4 states are represented by the combination of two binary communication signals. The receiver also generates appropriate motions according to the received sequential two signals. It can be said that 2bit serial communication emerged by RL of communication in a noisy environment.

5. Discussions, conclusions and future works

In this chapter, the author has advocated the introduction of parallel and flexible learning system and also the priority on optimization rather than understanding. As a concrete framework, it has been proposed that a neural network (NN) connects from sensors to actuators, and reinforcement learning (RL) trains the NN. The author hopes that through the above examples, the readers have felt a difference from the general approaches in conventional robotics research. From the viewpoint of the ratio of acquired functions vs. the prior knowledge, it is more than expected that even though only a scalar signal as reward and punishment is given except for sensor signals, various functions emerge flexibly, purposively and in harmony by the marriage of RL and a NN. Especially, in the communication-learning task, the gap between the reward for meeting and the establishment of two-bit serial communication is very interesting.

On the other hand, it is also true that the acquired abilities of robots or agents in the above examples have not yet reached the level to compare to human abilities and are insufficient from what the author claimed magnificently in the former part of this chapter. When we, humans, solve some difficult problem, we try to solve it by thinking logically, and a good idea often comes up suddenly. To realize such kind of intelligent and effective processing, an additional breakthrough may have to be waited for. However, the author still believes the proposed approach, which are introduction of parallel and flexible learning system and priority on optimization rather than understanding, is not wrong as a fundamental base to realize higher functions avoiding “Frame Problem” or “Symbol Grounding Problem”.

The tallest wall that the author can see towards the higher functions is “symbol emergence” and “logical thinking”. If it is assumed that symbols are equal to discrete representations, as shown in the communication-learning example(42), noise tolerance can be the necessity for the symbol emergence especially when communication is considered. The result that the two sequential signals are used to represent a piece of information makes the author feel the extensibility to the emergence of word or language. It is not impossible to consider that logical thinking is internalized communication, in other words, the communication from oneself to oneself, and logical thinking may not emerge before the communication with others is learned. If so, the noise tolerance can be the origin of symbol emergence.

Logical thinking has an important role; that is problem solving on a very high-level abstraction space. But, the author cannot imagine that from the necessity of such problem solving, logical thinking emerges. However, it is possible that forming of fixed-point attractors or chaos dynamics promotes symbol emergence. The dynamics makes only small number of isolated internal representations more stable, and on the contrary, makes the other most representations less stable. The author has suggested that fixed-point attractors are formed adaptively and purposively in a recurrent neural network (RNN) through learning(43). Tani’s group has shown some pioneering works in which dynamics of RNN is introduced in robotics research(44).

By introducing a RNN, the function of memory and state evaluation considering past state can be dealt with. However, the author feels the necessity to think more about the time axis in RL. For example, stochastic motion or action selection is usually done for exploration at each step in RL, but exploration should be designed considering the time axis. Temporal abstraction(45), improvement of the learning ability of RNN, and more flexible operation of time axis also should be considered more. One idea the author think now is that the state is not periodically sampled, but more event-based framework is introduced.

Recently, many works suggest that even newborn baby has much talent(46)(47), and it is a well-known fact that a horse can stand up soon after its birth. Furthermore, nowadays, the author has felt the limitation of the "learning from scratch" approach to develop higher functions. At least, our brain does not have a flat structure, but is structured, and the artificial neuron model used here is far simpler than our biological real neuron. The requirement of some modularized structure may be suggested to achieve coexistence of multiple dynamics such as talking while walking. Probably, initial knowledge should be introduced considering that it harms the flexibility as little as possible. The initial knowledge includes initial structure and connection weights, and also constraint of learning process.

Next, let us discuss the biological aspect. It is often said that RL is employed in the basal ganglia(48) and it is also said that the basal ganglia contributes action selection or decision making. Along the idea of function emergence by the couple of RL and a NN, author thinks that RL is highly possible to work in wider area including the frontal lobe, though the coexistence with another learning is not denied. Functions in the brain should be purposive, and RL can realize purposive function emergence. As for the NNs, the difference between artificial NN and biological NN was as mentioned. The author knows there are many negative opinions to Error Back Propagation (BP) learning when it is considered whether our brain employs BP or not. However, effective and harmonious learning and the autonomous and purposive division of roles in the network deeply thank to BP, in the authors' works. Furthermore, the author thinks that the studies of neurotrophic factors such as NGF(49) suggest us the existence of some backward signal flow for learning in our brain.

In the use of NN in RL, a difficulty has been pointed out(50), but the author does not think it so difficult if you prepare input signals in a local representation. However, when a RNN is used, it is a little bit difficult to adjust learning rate and bias to each neuron. Usually, learning rate should be small for feedback connections, and the bias should be 0 for the neurons in a loop. Setting self-feedback connection weights to be 4.0 often makes learning of memory easy and stable when a sigmoid function whose value range is from -0.5 to 0.5 or 0.0 to 1.0 is used as the output function. The symmetrical value range such as from -0.5 to 0.5 is good for stable learning. BPTT and RTRL are popular learning algorithms for RNNs, but are not practical from the viewpoint of computational cost and required memory capacity even though it is implemented in a parallel hardware. More practical learning algorithm with $O(n^2)$ computational cost and $O(n^2)$ required memory capacity is waited for(51)(52).

The author also thinks that towards realization of higher functions, "growing and developing from a simple form to a complicated one" should be the basic approach on behalf of "functional modularization". NNs are very flexible and very suitable for the approach. There is also a possibility that a NN structure flexibly and purposively changes according to learning(53)(54). However, it must be a difficult problem to design the growth and development. Introduction of evolutionary process may be inevitable.

Finally, the author would like to propose the necessity of supervision and discussion about this kind of research, considering the possibility of the emergence of out-of-control robots or systems(55).

6. Acknowledgement

At first, the author deeply thanks to all the students who tackled or are tackling to this impenetrable opponent. The AIBO task was done by Mr. Tomohiko Kawano, the memory task was done by Mr. Hiroki Utsunomiya, and the prediction task and exploration task are done by Mr. Kenta Goto. The author would like to express his gratitude to Prof. Yosuke Ito and Mr. Hironori Kakiuchi in Naruto University of Education for permitting us to use their AIBO walking software. The author also would like to express his appreciation to Prof. Yoich Okabe, Prof. Koji Ito, and Prof. Richard S. Sutton for giving him a chance to do a part of the above works and fruitful discussion. The author will never forget the exciting discussions with Prof. Sutton or Prof. Juergen Schmidhuber. This research has been supported for more than 10 years by MITI or JSPS Grant-in-Aid for Scientific Research and Research for the Future Program. Sorry for introducing only the most recent one, that is Grant-in-Aid for Scientific Research #19300070.

7. References

[1] Shibata, K. (2009). Emergence of Intelligence by Reinforcement Learning and a Neural Network, *Journal of SICE (Keisoku to Seigyo)*, Vol. 48, No. 1, pp. 106-111 (In Japanese)

[2] Shibata, K. & Kawano, T. (2009). Acquisition of Flexible Image Recognition by Coupling of Reinforcement Learning and a Neural Network, *SICE Journal of Control, Measurement, and System Integration (JCMSI)*, Vol. 2, No. 2, pp. 122-129

[3] McCarthy, J. & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence, *Machine Intelligence*, Vol. 4, pp. 463-502

[4] Dennett, D. (1984). Cognitive Wheels: The Frame Problem of AI, *The Philosophy of Artificial Intelligence*, Margaret A. Boden, Oxford University Press, pp. 147-170

[5] Tsumoto, T. (1986). *Brain and Development*, Asakura Publishing Co. Ltd. (in Japanese)

[6] Takatsuru, Y. et al. (2009). Neuronal Circuit Remodeling in the Contralateral Cortical Hemisphere during Functional Recovery from Cerebral Infraction, *J. Neuroscience*, Vol. 29, No. 30, pp. 10081-10086

[7] Johansson, P., Hall, L., Sikstrom, S. & Olsson, A. (2005). Failure to Detect Mismatches Between Intention and Outcome in a Simple Decision Task, *Science*, Vol. 310, pp. 116-119

[8] Brooks, R. A. (1991). Intelligence Without Representation. *Artificial Intelligence*, Vol. 47, pp.139-159

[9] Kawato, M. (1999). Internal Models for Motor Control and Trajectory Planning, *Current Opinion in Neurobiology*, Vol. 9, pp. 718-727

[10] Rumelhart, D. E. et al. (1986). Learning Internal Representation by Error Propagation, *Parallel Distributed Processing*, MIT Press, Vol. 1, pp. 318-364

[11] Irie, B. & Kawato, M. (1991). Acquisition of Internal Representation by Multilayered Perceptrons, *Electronics and Communications in Japan*, Vol. 74, pp. 112-118

[12] Saul, L. & Roweis, S. (2003). Think globally, fit locally: Unsupervised learning of nonlinear manifolds, *Journal of Machine Learning Research*, Vol. 4, pp. 119-155

[13] Littman, M.L., Sutton, R.S. & Singh, S. (2002). Predictive Representations of State, *In Advances in Neural Information Processing Systems*, Vol. 14, MIT Press, pp. 1555-1561

[14] Sutton, R. S., Rafols, E. J. & Koop, A. (2006). Temporal abstraction in temporal-difference networks, *Advances in Neural Information Processing Systems*, Vol. 18, pp. 1313-1320

[15] Bowling, M., Ghodsi, A. & Wilkinson, D. (2005). Action respecting embedding *Proc. of the 22nd Int'l Conf. on Machine Learning*, pp. 65-72

[16] Sperry, R.W. (1968). Hemisphere Deconnection and Unity in Conscious Awareness, *American Psy- chologist*, Vol. 23, pp. 723-733

[17] Harnad, S. (1990). Symbol Grounding Problem, Phisica D, Vol. 42, pp. 335-346

[18] Sutton, R. S. & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*, A Bradford Book, The MIT Press

[19] Anderson, C. W. (1987). Strategy learning with multilayer connectionist representations, *Proc. of the 4th Int'l Workshop on Machine Learning*, pp. 103-114

[20] Tesauro, G. J. (1992). TD-Gammon, a self-teaching backgammon program, achieves master-level play, *Neural Computation*, Vol. 6, No. 2, pp. 215-219

[21] Boyan, J. A. & Moore, A. W. (1995). Generalization in Reinforcement Learning, *Advances in Neural Information Processing Systems*, Vol. 7, The MIT Press, pp. 370-376

[22] Moody, J. & Darken, C. J. (1989). Fast Learning in Networks of Locally-tuned Processing Units, *Neural Computation*, Vol. 1, pp. 281-294

[23] Morimoto, J. & Doya, K. (2001). Acquisition of stand-up behavior by a real robot using hierarchical reinforcement learning, *Robotics and Autonomous Systems*, Vol. 36, No. 1, pp. 37-51

[24] Possio, T. & Girosi, F. (1990). Networks for Approximation and Learning, *Proc. of the IEEE*, Vol. 78, No. 9, pp. 1481-1497

[25] Albus, J. S. (1981). *Brain, Behavior, and Robotics*, Byte Books, Chapter 6, pp. 139-179

[26] Sutton, R. S. (1996). Generalization in reinforcement learning: Successful examples using sparse coarse coding, *Advances in Neural Information Processing Systems*, Vol. 8, MIT Press, pp. 1038-1044

[27] Shibata, K. & Ito, K. (2007). Adaptive Space Reconstruction on Hidden Layer and Knowledge Transfer based on Hidden-level Generalization in Layered Neural Networks, *Trans. SICE*, Vol. 43, No.1, pp. 54-63 (in Japanese).

[28] Shibata, K. & Ito, K. (1998). Reconstruction of Visual Sensory Space on the Hidden Layer in a Layered Neural Networks, *Proc. ICONIP '98*, Vol. 1, pp. 405-408

[29] Shibata, K., Sugisaka, M. & Ito, K. (2001). Fast and Stable Learning in Direct-Vision-Based Reinforcement Learning, *Proc. of AROB (Int'l Sympo. on Artificial Life and Robotics) 6th*, Vol. 1, pp. 200-203

[30] Shibata, K. and Ito, K. (1999). Gauss-Sigmoid Neural Network, *Proc. of IJCNN (Int'l Joint Conf. on Neural Networks) '99*, #747

[31] Schmidhuber, J. (2002). Exploring the Predictable, *Advances in Evolutionary Computing*, Springer, pp. 579-612

[32] Tani, J. (2003). Learning to Generate Articulated Behavior Through the Bottom-up and the Top-down Interaction Processes, *Neural Networks*, Vol. 16, No. 1, pp. 11-23

[33] Oudeyer, P. -Y., Kaplan, F. & Hafner, V. V. (2007). Intrinsic Motivation Systems for Autonomous Mental Development, *IEEE Trans. on Evolutionary Computation*, Vol. 11, No. 1, pp. 265-286

[34] McCracken, P. & Bowling, M. (2006). Online Discovery and Learning of Predictive State Representations, *Advances in Neural Information Processing Systems*, Vol. 18 , pp. 875-882

[35] Goto, K. & Shibata, K. (2010). Emergence of Prediction by Reinforcement Learning Using a Recurrent Neural Network, *J. of Robotics*, Vol. 2010, Article ID 437654

[36] Barto, A. G. et al. (1983). Neuronlike adaptive elements can solve difficult learning control problems, *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 13, No. 5, pp. 834-846

[37] Watkins, C. J. C. H. (1989). Learning from Delayed Rewards, *PhD thesis*, Cambridge University, Cambridge, England

[38] Shibata, K. & Kawano, T. (2009). Learning of Action Generation from Raw Camera Images in a Real-World-like Environment by Simple Coupling of Reinforcement Learning and a Neural Network, *Advances in Neuro-Information Processing*, LNCS, Vol. 5506, pp. 755-762

[39] Utsunomiya, H. & Shibata, K. (2009). Contextual Behavior and Internal Representations Acquired by Reinforcement Learning with a Recurrent Neural Network in a Continuous State and Action Space Task, *Advances in Neuro-Information Processing*, LNCS, Vol. 5506, pp. 755-762

[40] Goto, K. & Shibata, K. (2010). Acquisition of Deterministic Exploration and Purposive Memory through Reinforcement Learning with a Recurrent Neural Network, *Proc. of SICE Annual Conf. 2010*

[41] Shibata, K. (2006). Spatial Abstraction and Knowledge Transfer in Reinforcement Learning Using a Multi-Layer Neural Network, *Proc. of Fifth Int'l Conf. on Development and Learning*

[42] Shibata, K. (2005). Discretization of Series of Communication Signals in Noisy Environment by Reinforcement Learning, *Adaptive and Natural Computing Algorithms, Proc. of ICANNGA'05*, pp. 486-489

[43] Shibata, K. & Sugisaka, M. (2002). Dynamics of a Recurrent Neural Network Acquired through the Learning of a Context-based Attention Task, *Proc. of AROB (Int'l Sympo. on Artificial Life and Robotics) 7th*, pp. 152-155

[44] Yamashita, Y. & Tanni, J. (2008). Emergence of Functional Hierarchy in a Multiple Timescale Neural Network Model: a Humanoid Robot Experiment, *PLoS Comupt. Biol.*, Vol. 4, No. 11, e1000220

[45] Shibata, K. (2006). Learning of Deterministic Exploration and Temporal Abstraction in Reinforcement Learning, *Proc. of SICE-ICCAS (SICE-ICASE Int'l Joint Conf.)*, pp. 4569-4574

[46] Bremner, J. G. (1994). *Infancy*, Blackwell Publishers limited

[47] Meltzoff, A. N. & Moore, M. K. (1994). Imitation, memory, and the representation of persons, *Infant Behavior and Development*, Vol. 17, pp. 83-99

[48] Schultz, W. et al. (1992). Neuronal Activity in Monkey Ventral Striatum Related to the Expectation of Reward, *J. of Neuroscience*, Vol. 12, No. 12, pp.4595-4610

[49] Brunso-Bechtold, J.K. and Hamburger, V. (1979). Retrograde transport of nerve growth factor in chicken embryo, *Proc. Natl. Acad. Sci. USA.*, No.76, pp. 1494-1496

[50] Sutton, R. S. (2004). Reinforcement Learning FAQ, http://webdocs.cs.ualberta.ca/ sutton/RL-FAQ.html

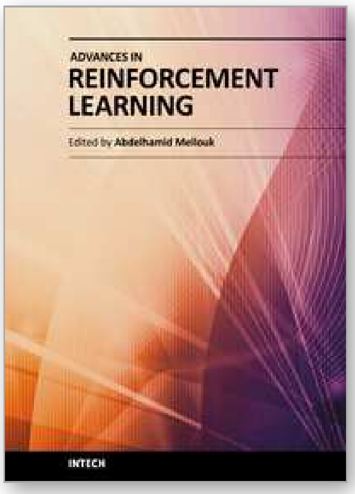
[51] Shibata, K., Ito, K. & Okabe, Y. (1985). Simple Learning Algorithm for Recurrent Networks to Realize Short-Term Memories, *Proc. of IJCNN (Int'l Joint Conf. on Neural Networks) '98*, pp. 2367-2372

[52] Samsudin, M. F. B., Hirose, T. and Shibata, K. (2007). Practical Recurrent Learning (PRL) in the Discrete Time Domain, *Neural Information Processing of Lecture Notes in Computer Science*, Vol. 4984, pp. 228-237

[53] Kurino, R., Sugisaka, M. & Shibata, K. (2003). Growing Neural Network for Acuqisition of 2-layer Structure, *Proc. of IJCNN (Int'l Joint Conf. on Neural Networks) '03*, pp. 2512-2518

[54] Kurino, K., Sugisaka, M. & Shibata, K. (2004). Growing Neural Network with Hidden Neurons, *Proc. of The 9th AROB (Int'l Sympo. on Artificial Life and Robotics)*, Vol. 1, pp. 144-147

[55] Joy, B. (2000). Why the Future doesn't Need Us, *WIRED*, Issue 8.04



ISBN 978-953-307-369-9
Hard cover, 470 pages
Publisher InTech
Published online 14, January, 2011
Published in print edition January, 2011

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821