Operant Conditioning Learning Model in the Bionic Experiment

Jing Huang[1, 2,a], Xiaogang Ruan[1], Qingwu Fan[2] and Xiaoping Zhang[1]

[1] Institute of Artificial Intelligence and Robotics, Beijing University of Technology, Beijing, China

[2]Pilot College, Beijing University of Technology, Beijing, China

[a]aiandrobot@163.com

Keywords: Operant Conditioning Learning Model, Skinner Rat Experiment, Autonomous Learning, Cognitive Robotics

Abstract. A learning model based on the operant conditioning mechanism (*OCLM*) is presented in this paper to deal with the autonomous learning problem in cognitive robotics. The model is described by 9 elements, including the space set, the action set, the bionic learning function and the system entropy etc. To describe the learning mechanism which is the core of the model, a new notion "*negative ideal degree*"(*NID*) is defined. We also prove the convergence of *OCLM* to indicate that the model is a self-organization system. *OCLM* has been applied to simulating the Skinner rat experiment. The results show that this model can well simulate the animal's operant conditioning behavior, acquire the cognitive skills through the interaction with the environment and achieve self-learning and self-adaptability.

Introduction

Since it was born in the nineteen fifties, the theory of cognitive psychology has profoundly influenced the study of artificial intelligence, robotics and related fields. According to the embodied cognition approach of cognitive psychology[1], cognition is situated and the cognitive activities always occur in a certain objective real environment. The environment is part of the cognitive system, and cognition is realized through the frequent interaction between the cognitive subject and the environment. In fact, the theory is the same with the design philosophy of cognitive robotics. Brooks[2] designed a robust hierarchical control structure to control the mobile robot in 1985. He[3] described the design idea of the control structure as direct interaction with the environment through the perception and action. J. Weng[4] developed this theory, put forward the concept "*autonomous mental development*", and emphasized that robot can develop the intelligence autonomously through the continuous interaction with the environment.

The relationship between the action and feedback from environment in learning process has been discussed by many researchers, one of whom is B.F. Skinner. He for the first time put forward the operant conditioning principles in his book *The Behavior of Organisms: an experimental analysis*[5] in 1938. It clearly described how actions would be influenced by their consequences. Since the middle of nineteen nineties, many researchers have carried out the study on the application of operant conditioning to robotics. Gaudiano[6,7] from Boston University of the United States proposed a neural network model based on operant conditioning and classical conditioning and applied it to the obstacle avoidance for wheeled robots. At the same time, Touretzky[8] from Carnegie Mellon University put forward a calculation model about the operant conditioning to compensate the simplicity of the reinforcement learning, then applied it to RWI B21 real robot. The model was verified by reproducing animal learning experiments. In 2013, Manoonpong[9] modeled classical conditioning as correlation-based learning and operant conditioning as reinforcement learning or reward-based learning, and proposed a dual learner system by combining the two models together. The model performance was evaluated by simulations of a cart-pole system as a dynamic motion control problem and a mobile robot system as a goal-directed behavior control problem. Although multiple models for operant conditioning have been presented above, a unified and formulated computation model is still absent. J.X. Cai et al. from Beijing University of Technology has presented

an automata based on operant conditioning to resolve the self-balance problem in two-wheeled robots since 2010[10,11]. The model is formulated and was proved to be efficient, but there is room to improve it. For example, the orientation function as the core part of the automata is inconsistent with the counterpart in biology in definition. The proof for the convergence of the automata isn't enough or rigorous.

In order to solve the above mentioned problems, this paper presents an operant conditioning learning model (*OCLM*),whose validity is showed by reproducing the classical Skinner rat experiment. The rat in the simulation program can develop cognitive ability through interaction with the environment, which completely reflects the embodied cognition view.

Skinner's operant conditioning theory and experiment

Operant conditioning is one of the most famous theories of B.F.Skinner. He drawed on Pavlov's reinforcement concept, and innovated its connotation. Behind the theory it is the phenomenon that human beings or animals have the instinct to seek benefits and avoid disadvantages. Therefore, Skinner divided stimuli into positive reinforcement and negative reinforcement. Positive reinforcement will increase the probability of the action which produce the stimulus, while the negative reinforcement will decrease the probability of the action. At the same time, the positive reinforcement is more effective than the negative reinforcement in learning process. In a word, actions will produce stimuli, then stimuli will influence the probability of actions as feedback, and the impacts of different stimuli on actions are different. That is the main content of Skinner's operant conditioning theory.

Skinner's operant conditioning theory reflects the self-adaptation of agents including human beings and animals to the environment. The process during which operant conditioning mechanism is established is also an autonomous-learning process during which cognitive ability is gradually developed. The abstract cognitive model in the frame of Skinner operant conditioning is shown in Fig. 1.In Fig. 1 the line representing negative reinforcement is thinner than the one reprenting positive reinforcement to indicate positive reinforcement is more influential in action choosing than negative reinforcement.

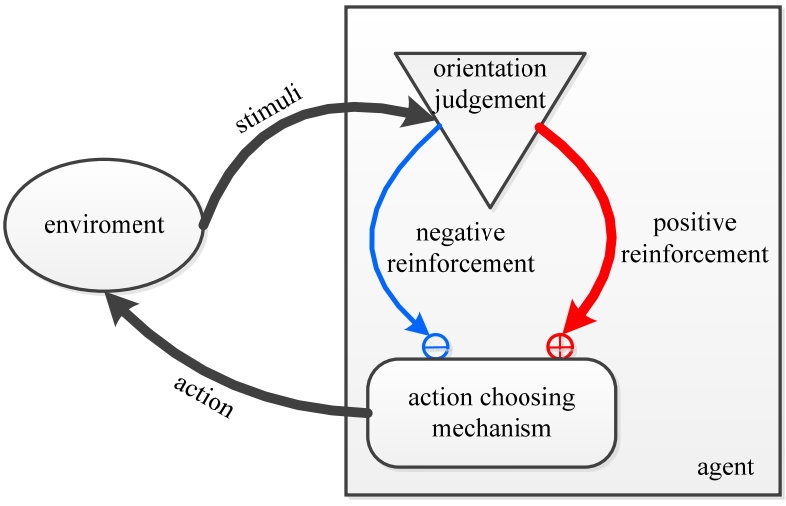

Fig. 1 Cognitive model in the frame of Skinner opearant conditioning

In order to study and prove the operant conditioning theory, Skinner had designed a lot of animal experiments, among which Skinner rat experiment is the most famous one. He placed a rat in *Skinner box* , in which there is a bar or lever that the rat can press to receive food and water, and a device that records the organism's responses. In the beginning, the rat randomly pressed the lever and got food. After several times of trials, it learned the rule and the pressing action became frequently, which indicated that the rat had already build up operant conditioning. The equipment of the experiment is shown in Fig. 2.

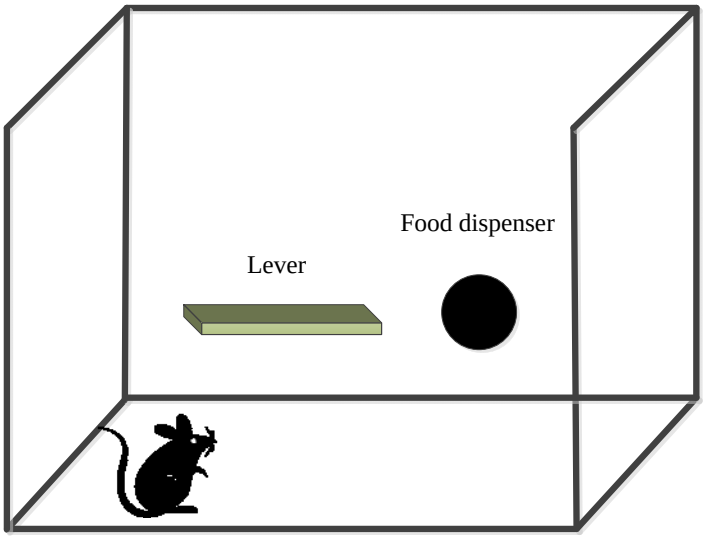

Fig. 2 Skinner rat experiment equipment

Operant Conditioning Learning Model

The *Operant Conditioning Learning Model (OCLM)* is defined by nine elements:
$OCLM =< t, S, A, P, f, \varepsilon, \delta, L, H >$,and it is defined as follows：

1.*t*:time parameter of the *OCLM*, and also represents the model iterative times, $t = \{t_i \mid i = 0,1,\cdots, n_t\}$, t_0 represents initial establishment time of the model.

2.*s*: the state space of the *OCLM*, $S = \{s_i \mid i = 1,\cdots, n_s\}, s_i$ represents the *i-th* state in the state space, and n_s is the size of the state space which means the number of states.

3.*A*: the set of actions in the *OCLM*, $A = \{a_k \mid k = 1,2,\cdots, n_a\}$, a_k represents the *k-th* action and n_a is the numbers of actions.

4.*P*: the probability distribution of *OCLM*, $P : S \times A \to P = \{p_{ik} \mid i = 1,2,\cdots, n_s, k = 1,2,\cdots, n_a\}$, in which $p_{ik} = p(a_k \mid s_i)$ represents the probility of choosing $a_k \in A$ when the rat is in the state $s_i \in S$.

P can also be written as the vector format $P = \{P_1, P_2, ..., P_{n_s}\}$, in which $P_i = \{p_{i1}, p_{i2}, \ldots, p_{ir}\}$ means the probility vector for the *i-th* state and *r* means the number of possible actions at that state. Obviously, $\sum_{k=1}^{r} p_{ik} = 1$ and $0 \le p_{ik} \le 1$.

5.*f*:the state transfer function in *OCLM*, $f : S \times A\big|_P \to S$,means that the state $s_i \in S$ will be transferred from to $s_j \in S$ after the rat chooses the $a_k \in A$ with the probility of $p_{ik} \in P$.

6. ε :the negative ideal degree of each state. The *negative ideal degree* is presented in this paper to caculate the orientation function. The definition is as follows:

Definition 1: *Negative ideal degree (NID)*

NID is denoted as $\varepsilon = \varepsilon(S) = \{\varepsilon(s_i) \mid i = 1,2,\cdots, n_s\} \in R$ and it indicates how negative ideal the state is. The bigger value of *NID*, the farther away from ideal state for state s_i. Its caculation method is:

a.If the output of the system is quantitive: denote the ideal output influenced by state s_i as x^* and the actual output as x , then $\varepsilon = f_1(x - x^*)$, in which f_1 is the monotone increasing function about the absolute value of $(x - x^*)$.

b.If the output of the system is not quantitive: referring to how to define fuzzy sets in fuzzy mathematics, the state space can be divided into *N* sets: $S = (Set_1 \bigcup Set_2 \bigcup \cdots \bigcup Set_N)$ .There are no

intersections among the sets. Suppose $s_i \in Set_m, s_j \in Set_n$, the relationship of *NIDs* between the two states is:

$$\begin{cases} \varepsilon(s_i) = \varepsilon(s_j) & m = n \\ \varepsilon(s_i) \neq \varepsilon(s_j) & m \neq n. \end{cases} \tag{1}$$

And the value of *NID* in each set can be defined according to the details of the problem.

7. δ :the orientation funciton in the *OCLM*. The orientation function $\delta = \delta(S, A) = \{\delta_{ik} \mid i = 1, 2, \cdots, n_s; k = 1, 2, \cdots, n_a\}$ is introduced to simulate the biological orientation in nature. Symbol δ_{ik} means the performance changes of the system after choosing $a_k \in A$ at state $s_i \in S$. To keep identical with the biological orientation in meaning, if $\delta > 0$, the orientation is positive and indicates that the performance of the system tends to become better, while if $\delta < 0$, the orientation is negative and indicates that the performance tends to become worse. If $\delta = 0$, it indicates that the performance has no change. Based on the definiton of *NID*, the orientation function δ is defined as follows:

Definition 2: Orientation function δ

Suppose action a_k is excuted in the state s_i and then the state is tranferred to s_j, δ_{ik} is defined as follows：

$$\delta_{ik} = \delta(\Delta\varepsilon_{ij}) \begin{cases} >0, & \Delta\varepsilon_{ij} < 0 \\ = 0\ , & \Delta\varepsilon_{ij} = 0 \\ <0, & \Delta\varepsilon_{ij} > 0\ . \end{cases} \tag{2}$$

in which $\Delta\varepsilon_{ij} = \varepsilon(s_j) - \varepsilon(s_i)$. The orientation function is continous in the defined interval and is monotone decreasing function about $\Delta\varepsilon_{ij}$ and its absolute value is monotone increasing about the absolute value of $\Delta\varepsilon_{ij}$.

Here we provide the explanation for the definition as follows. If $\Delta\varepsilon_{ij} > 0$, it means the negative ideal degree is increasing, i.e. the system performance becomes worse, therefore the orientation function $\delta < 0$. Moreover, the bigger $\Delta\varepsilon_{ij}$, the smaller δ. On the contrary, if $\Delta\varepsilon_{ij} < 0$, the negative ideal degree is decreasing, which indicates the system performance becomes better, so the orientation function $\delta > 0$. Similarly, the bigger $\Delta\varepsilon_{ij}$, the smaller δ. Finally,if $\Delta\varepsilon_{ij} = 0$, the negative ideal degree has no change, indicating the system performance doesn't change, so the orientation function $\delta = 0$.

Base on the orientation function, *positive reinforcement* and *negative reinforcement* in operant conditioning can be defined as follows:

Definition 3 Positive reinforcement and negative reinforcement

Suppose *OCLM* executes a_k in the state of s_m and experiences the stimulus θ from the environment, then the state is transferred to s_n. If $\delta_{mk} = \delta(\Delta\varepsilon_{mn}) > 0$， then θ is called *positive reinforcement* and is denoted as θ^+. On the contrary, if $\delta_{mk} = \delta(\Delta\varepsilon_{mn}) < 0$, then θ is called *negative reinforcement* and is denoted as θ^-. If $\delta_{mk} = \delta(\Delta\varepsilon_{mn}) = 0$， then θ is called *neutral stimulus* and is denoted as θ^N .

8.*L*: the learing mechanism of the *OCLM*, $L : P(t) \to P(t+1)$ ($P(t)$ is the probability distribution at time t). *L* will adjust the probalility distribution according to operant coditioning,which means that the possiblity of the actions causing the positive reinforcement will be increased while the probability

of the actions causing the negative reinforcement will be decreased. Suppose *OCLM* executes a_k in the state of s_m and experiences the stimulus θ from the environment, then the state is transferred to s_n:

If $\theta = \theta^+$, then:

$$L: \begin{cases} p_{mk}(t+1) = p_{mk}(t) + \dfrac{1 - p_{mk}(t)}{1 + \exp\left(-\eta_1 \delta_{mk} \cdot t\right)}, & a(t) = a_k \\ p_{mk'}(t+1) = p_{mk'}(t) - \dfrac{1 - p_{mk}(t)}{1 + \exp\left(-\eta_1 \delta_{mk} \cdot t\right)} \cdot \dfrac{1}{n_a - 1}, & a(t) \neq a_k. \end{cases} \tag{3}$$

If $\theta = \theta^-$, then:

$$L: \begin{cases} p_{mk}(t+1) = p_{mk}(t) - \dfrac{p_{mk}(t)}{1 + \exp\left(\eta_2 \delta_{mk} \cdot t\right)}, & a(t) = a_k \\ p_{mk'}(t+1) = p_{mk'}(t) + \dfrac{p_{mk}(t)}{1 + \exp\left(\eta_2 \delta_{mk} \cdot t\right)} \cdot \dfrac{1}{n_a - 1}, & a(t) \neq a_k. \end{cases} \tag{4}$$

If $\theta = \theta^N$, then the probability keeps unchanged, that is:

$$L: \begin{cases} p_{mk}(t+1) = p_{mk}(t), & a(t) = a_k \\ p_{mk'}(t+1) = p_{mk'}(t), & a(t) \neq a_k. \end{cases} \tag{5}$$

in which $p_{mk}(t)$ is the probability that *OCLM* execute a_k at the state of s_m at time t. Obviously,the ajustment to the probability should always satisty $\sum_{k=1}^{n_a} p_{ik} = 1$ and $0 \le p_{ik}(t) \le 1$.

η_1, η_2 here is the learning rate and $\eta_1 > \eta_2 > 0$. The two constants are used for adjusting the changing speed of the probability. To reflect the theory that positive reinforcement is effective than the negative reinforcement, set $\eta_1 > \eta_2$ in order to fast the learning speed for the positive reinforcement.

9.*H*: the system entropy in the *OCLM*. The proposed *OCLM* attains cognitive ability ultimately by interacting with the environment, and the cognitive process itself is a self-adaptive process which is based on the system's self-organization. When the system changes from disorder to order, it realizes the self-adaptivity with the process of self-organization. The system entropy *H* is used to describe the self-organization of the model and then explains the adaptivity of the model. To calculate it, *state entropy HS* is presented first and is defined as follows:

Definition 4 State entropy HS

State entropy $HS(t) = \{HS_i \mid i = 1, 2, \cdots, n_s\}$ is used to reflect the chaos degree for any state s_i at time t, and $HS_i = HS\left(A \mid s_i\right) = -\sum_{k=1}^{n_a} p(a_k \mid s_i) \log_2 p(a_k \mid s_i)$.

Definition 5 System entropy H

System Entropy $H(t)$ is used to describe the chaos degree of the model at time *t*. It is the mathematical expectation of the state entropy $HS(t)$, that is:

$$H(t) = -\sum_{i=1}^{n_s} p(s_i) \sum_{k=1}^{n_a} p(a_k \mid s_i) \log_2 p(a_k \mid s_i). \tag{6}$$

Obviously, the system entropy $H(t)$ is inversely proportional to the self-organization degree of the system. The smaller $H(t)$, the higher the self-organization degree of the system.

In summary, the working principle of the *OCLM* can be described as follows: suppose at time *t*, the negative ideal degree of the state $s_i \in S$ is ε_i and $a_k \in A$ is selected with the probability $p_{ik} \in P$. Then the state will be transferred to state s_j. After receiving the stimulus from the environment for this action, the negative ideal degree is changed to ε_j. Then the orientation function $\delta(\Delta\varepsilon) = \delta(\varepsilon_j - \varepsilon_i)$ can be caculated. Based on it the probability distribution function P can be adjusted according to the learning mechanism L. After multiple rounds of iterative calculation, the model can acquire the optimal action, forming operant conditioning and reaching the minimum of the system entropy H which means self-organization has been realized. The flow diagram of the algorithm is illustrated in fig. 3.

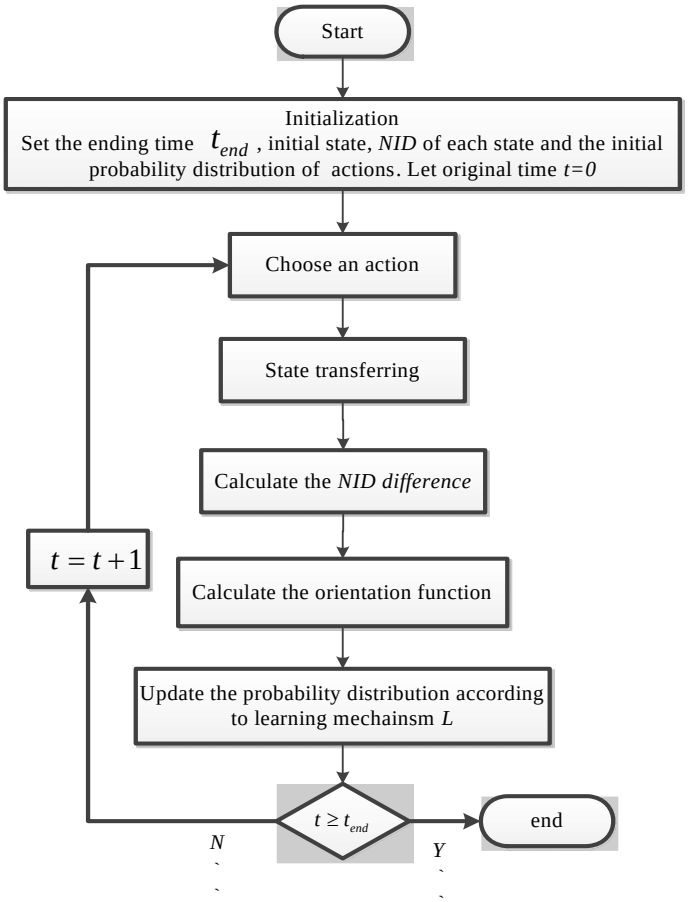

Fig. 3 Working algorithm flow chart of *OCLM*

Convergence Analysis

A self-learning model can achieve self-adaptibility to the environment and the self-organization. To prove that the *OCLM* is such a model, the convergence of the model is analyzed as follows:

Theorem 1

For $OCLM =< t, S, A, P, f, \varepsilon, \delta, L, H >$:

$$\lim_{t\to\infty} p_{ik}(a_k(t)|s_i(t)) = 1, \tag{7}$$

$$\lim_{t\to\infty} p_{ik'}(a_{k'}(t)|s_i(t)) = 0 \quad (k' \neq k), \tag{8}$$

in which $a_k(t)$ means the action corresponding to the orientation function $\delta_{ik} > 0$ at time t, and $a_{k'}(t)$ means the action corresponding to the orientation function $\delta_{ik} \le 0$ at time t.

Proof of theorem 1

1. If *OCLM* at state s_i selects the action $a_k \in A$, $\delta_{ik} > 0$, then according to formula (3):

$$\Delta p_{ik}(t) = p_{ik}(t+1) - p_{ik}(t) = \frac{1 - p_{ik}(t)}{1 + \exp\left(-\eta_1 \delta_{ik} \cdot t\right)}. \tag{9}$$

$\because \eta_1, \delta_{ik} > 0$, then $\lim_{t\to\infty} \Delta p_{ik}(t) = 1 - p_{ik}(t)$, then $\lim_{t\to\infty} p_{ik}(t+1) = \lim_{t\to\infty}\left[p_{ik}(t) + \Delta p_{ik}(t)\right] = 1$. That is $\lim_{t\to\infty} p_{ik}(a_k(t)|s_i(t)) = 1$.

2. If *OCLM* at state s_i selects the action $a_{k'} \in A$, $\delta_{ik'} < 0$, $k' \ne k$, then according to formula (4):

$$\Delta p_{ik'}(t) = p_{ik'}(t+1) - p_{ik'}(t) = -\frac{p_{ik'}(t)}{1 + \exp\left(\eta_2 \delta_{ik'} \cdot t\right)}. \tag{10}$$

$\because \delta_{ik'} < 0, \eta_2 > 0$, then $\lim_{t\to\infty} \Delta p_{ik'}(t) = -p_{ik'}(t)$, then $\lim_{t\to\infty} p_{ik'}(t+1) = \lim_{t\to\infty}\left[p_{ik'}(t) + \Delta p_{ik'}(t)\right] = 0$, that is $\lim_{t\to\infty} p_{ik'}(a_{k'}(t)|s_i(t)) = 0 \ \ (k' \ne k)$.

3. If *OCLM* at state s_i selects the action $a_{k'} \in A$, $\delta_{ik'} = 0$, suppose $a_{k1}, a_{k2} \in A$, $\delta_{ik1} > 0, \delta_{ik2} < 0$, which represent 2 kinds of actions, then:

$$p_{ik'}(a_{k'}(t)|s_i(t)) = 1 - p_{ik1}(a_{k1}(t)|s_i(t)) - p_{ik2}(a_{k2}(t)|s_i(t)). \tag{11}$$

According to the conclusions proved above, we can derive that:

$$\begin{aligned} \lim_{t\to\infty} p_{ik'}(a_{k'}(t)|s_i(t)) &= \lim_{t\to\infty}[1 - p_{ik1}(a_{k1}(t)|s_i(t)) - p_{ik2}(a_{k2}(t)|s_i(t))] \\ &= 1 - \lim_{t\to\infty} p_{ik1}(a_{k1}(t)|s_i(t)) - \lim_{t\to\infty} p_{ik2}(a_{k2}(t)|s_i(t)) \\ &= 1 - 1 - 0 = 0. \end{aligned} \tag{12}$$

Theorem 2

For $OCLM = < t, S, A, P, f, \varepsilon, \delta, L, H >$, the system entropy H will be converged to a minimum When $t \to \infty$, i.e. $\lim_{t\to\infty} H(t) = H_{\min}$, in which $H_{\min}$ is a constant.

Proof of theorem 2

1. When $t = 0$: since the initial probability distribution is uniform, the probabilities of all actions in any state s_i are the same. According to the definition of state entropy, the state entropy HS_i will reach a maximum value. Therefore the system entropy $H(t=0) = -\sum_{i=1}^{n_s} p(s_i) \sum_{k=1}^{n_a} p(a_k \mid s_i) \log_2 p(a_k \mid s_i)$ will also reach its maximum value.

2. When $t \to \infty$, the state entropy for any state s_i in *OCLM* can be calculated as follows:

$$HS_i = HS(A \mid s_i) = -\left[\sum_{k=1}^{n_1} p(a_k \mid s_i) \log_2 p(a_k \mid s_i) + \sum_{k'=1, k' \ne k}^{n_2} p(a_{k'} \mid s_i) \log_2 p(a_{k'} \mid s_i)\right], \tag{13}$$

in which a_k indicates actions producing positive reinforcement while $a_{k'}$ indicates actions producing negative reinforcement and neutral stimulus. The symbol n_1 is the total number of a_k and n_2 is that of $a_{k'}$. Obviously, $n_1 + n_2 = n_a$.

According to the conclusions of *theorem 1* we can get: $\lim_{t\to\infty} HS_i(t) = 0$, then the system entropy $H(t) = -\sum_{i=1}^{n_s} p(s_i) HS_i \mid_t \to 0 \ \ (t \to \infty)$, i.e. $\lim_{t\to\infty} H(t) = H_{\min}$.

From the analysis above we can conclude that during the learning process the system entropy is continuously decreasing and approaches to 0, which means the self-organization degree is keeping increasing and the self-organization is finally realized.

Bionic Experiment

As proved above, the *OCLM* is a self-organization system and has the self-learning ability. This section will reproduce the classical *Skinner rat experiment* to verify its validity. In the *Skinner rat experiment*, the actions of the rat include: pressing the bar a_1 and not pressing the bar a_2, i.e. the action set $A = \{a_1, a_2\}$. The state set is $S = \{s_0, s_1\}$, in which s_0 means full, s_1 means hungry. Considering the specific circumstances in the experiment, set the *NID* of each state to be $\varepsilon = \varepsilon(S) = \{\varepsilon(s_0) = -1, \varepsilon(s_1) = 1\}$. The state transfer function $f : S \times A|_P \to S$ in the model can be expressed in Fig.4:

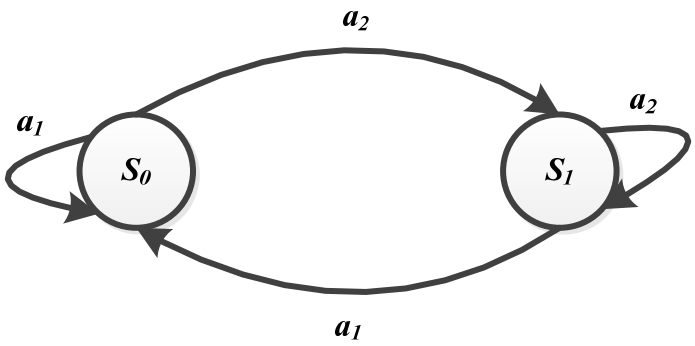

Fig. 4 State transfer diagram for *Skinner rat experiment*

For example, the state will be transferred to s_1 from s_0 after a_2 is executed and the rest can be explained in the same way.

Based on the definition requirements, the orientation function in this experiment is defined as follows:

$$\delta_{ik} = \delta(\Delta\varepsilon_{ij}) \begin{cases} \exp(1/\Delta\varepsilon_{ij}), & \Delta\varepsilon_{ij} < 0 \\ = 0\,, & \Delta\varepsilon_{ij} = 0 \\ -\exp(-1/\Delta\varepsilon_{ij}), & \Delta\varepsilon_{ij} > 0. \end{cases} \tag{14}$$

in which $\Delta\varepsilon_{ij} = \varepsilon_j - \varepsilon_i$, that is the negative ideal degree changes after the state transfer.

Let the initial state be s_0 , and the initial probability of actions be uniform, which means $p_{0k} = 0.5\ (k = 1,2)$, the positive reinforcement learning rate $\eta_1 = 0.2$ and the negative reinforcement learning rate $\eta_2 = 0.1$. The model can converge fast after about 20 times of learning.The results are shown in Fig. 5 and Fig. 6.

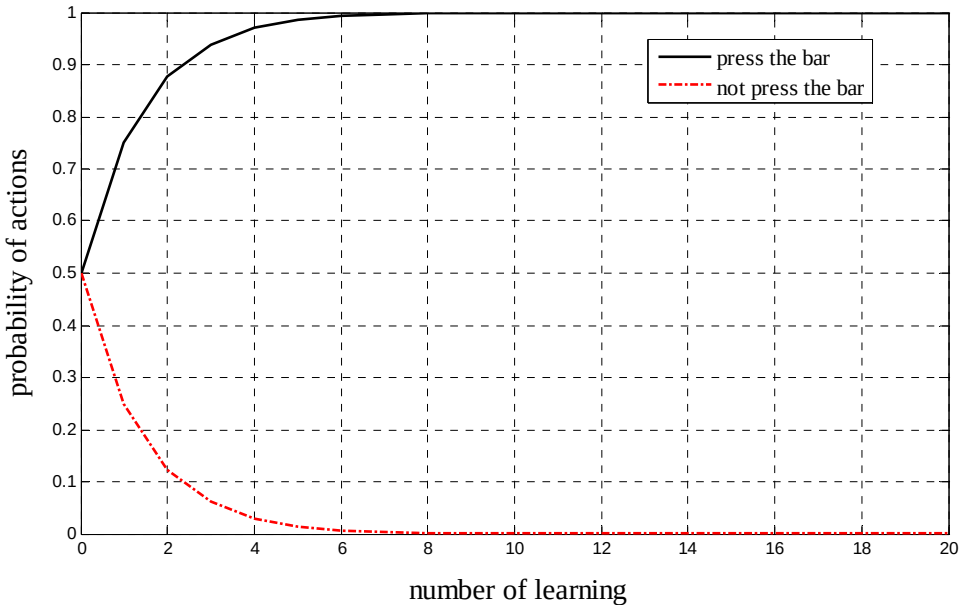

Fig.5 The change of probability

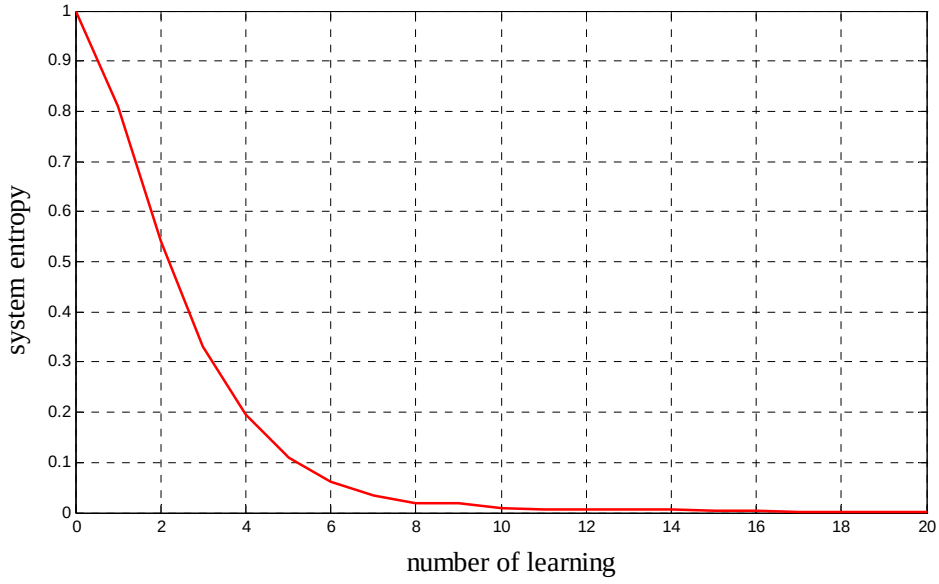

Fig.6 The change of system entropy

The experimental results above show that the probabilities for the two actions(pressing the bar or not) are the same at the beginning and are continuously changeing with the progress of learning according to the operant conditioning priciple. The probability to press the bar keeps increasing while the other one keeps decreasing. After 20 times learning, the probability to press the bar is far more than the other one and converges to 1, which indicates that the rat has learned to press the bar to get food and the operant conditioning has been formed. On the other hand, the state entropy of *OCLM* shown in Fig. 6 is continuously decreasing during the process and finally converges to 0, which

indicates that the model is going from disorder to order and the self-organization has been realized in the end.

From the analysis above we can see that the bionic experiment results are in accord with those of *Skinner rat experiment*, which proves that *OCLM* presented in this paper can simulate animal learning process and gradually increase the cognitive ability by interaction with the enviroment. The self-learning process itself is a self-organization process as well, during which self-adaptivity is the purpose and nature.

Conclusion

This paper presents a bionic autonomous learning model *OCLM*. It can be described with 9 elements, in which the design and definition of learning mechanism is the core. The learning mechanism is based on the Skinner operant conditioning mechanism and adjusts the action probability distribution of every state according to the principle that the probability of actions producing positive reinforcements increases and the one of actions producing negative reinforcements decreases. This paper also presents the concept of negative ideal degree*(NID)* to calculate the orientation function. Then the positive reinforcements and negative ones can be distinguished by the value of the orientation function. The convergence of *OCLM* is proved in this paper to explain the self-organization of the model.

To verify the validity of *OCLM*, *Skinner rat experiment* is reproduced and simulated. The results show that this model can well simulate the animal's operant conditioning behavior, acquire the cognitive skills through the interaction with the environment and achieve self-learning and self-adaptability. The experiment data also shows that *OCLM* is a self-organization system. This paper lays a foundation for further study on the cognitive model of robots.

Acknowledgements

This work was financially supported by National Natural Science Foundation of China (No.61075110), Key Project (No.KZ201210005001) of S&T Plan of Beijing Municipal Commission of Education, National Basic Research Program of China(973 Program) (2012CB720000), and Specialized Research Fund for the Doctoral Program of Higher Education (No.20101103110007).

References

[1] H.W. Li, J.Y. Xiao, Journal of Dialectics of Nature. Vol. 28(1) (2006), p. 29 (In Chinese)

[2] R. A. Brooks: IEEE J. Robotics and Automation. Vol. 2(1) (1986), p. 14

[3] R.A. Brooks: Artificial Intelligence. Vol. 47(1)(1991), p. 139

[4] J. Weng, J. Mcclelland, A. Pentland et al.: Science. Vol. 291(2001), p. 599

[5] B.F. Skinner: *The Behavior of organisms: an experimental analysis* (Appleton-Century Company, New York 1938)

[6] P. Gaudiano, C. Chang: IEEE International Symposium on Computational Intelligence in Robotics and Automation (1997), p. 13

[7] P. Gaudiano, E. Zalama, C. Chang and J. L. Coronado: *From Animals to Animats 4*(1996),p.373

[8] D.S. Touretzky and L.M. Saksida: Adaptive Behavior Vol.5 No.3-4 (1997), p.219

[9] P. MANOONPONG, C. KOLODZIEJSKI, F. WÖRGÖTTER and J. MORIMOTO: Advances in Complex Systems(2013), in press

[10] X.G. Ruan, J.X. Cai and L.Z. Dai: Journal of Beijing University of Technology Vol.36 No.8 (2010), p. 1025 (In Chinese)

[11] X.G. Ruan, L.Z. Dai, N.G. Yu and J.J. Yu: Control Theory & Applications Vol. 29 No.11(2012), p.1452(In Chinese)