Basal Ganglia Cognitive Behavioral Model Based on Operant Conditioning Reflex

Chen Jing* and Li Zongshuai

School of Information Technology Engineering, Tianjin University of Technology and Education, 300222 Tianjin, China

Aeronautical Automation College, Civil Aviation University of China, 300300 Tianjin, China

Abstract: Aiming at autonomous biological learning problems, applying the winner-takes-all (WTA) learning mechanism, using the principle of cortical-striatal synaptic modification principle in basal ganglia, based on the operant conditioning reflex theory, a basal ganglia behavior cognitive model is established. The proposed cognitive model is suitable for the cognitive learning of limited actions. The applications of the bionic learning model for Skinner's pigeons experiment were simulated. The experimental results show that through the cognitive behavioral model to simulate the gradual process of adaptive learning for Skinner pigeons, the animals showed a gradual process of adaptive learning. This research provides a reference for bionic cognitive model of agent.

Keywords: Basal ganglia, cognitive model, entropy, operant conditioning reflex, pigeon experiment.

1. INTRODUCTION

Skinner's Operant Conditioning Reflex (OCR) principle has been extensively used in the animal training, teaching and medical science. In recent years, OCR method has been used with respect to machine learning and control of robot, and a variety of experiments have been performed.

In 1988, from University of California, Bruce and others established an OCR-based learning model based on recursive learning in reinforcement learning algorithm from Klopf's assumption [1], and applied it in an upright and balancing robot. In 1995, Zalama *et al.* researched on the obstacle avoidance for robot using the OCR model [2]. Robots have learnt the behavior of obstacle avoidance at any position after a period of conditioning reflex learning and training, although in the beginning the robot was in an environment characterized by disorder. In 1997, Gaudiano *et al.* performed a physical experiment on a robot named Pioneer 1 using operant conditioning learning method through off-line learning, and then transplanted the well-trained weights into the control program, to get a better obstacle avoidance effect [3]. Björn Brembs [4, 5], from Germany, conducted a research on OCR in drosophila, and designed a simulation device for a bionic aircraft, and he pointed out that prediction is very important for the agent cognition, which directs to the next decision. In 2002, Zalama [6] *et al.* designed a neural network model for the reactive behavioral navigation of a mobile robot based on OCR theory. In 2005, Itoh [7] made a robot, named WE-4RII, that learnt the handshake behavior using the Hull theory based on OCR theory. In 2007, Tadahiro, from University of Tokyo, designed an OCR model using incremental acquisition of behaviors and signs based on a reinforcement learning schemata model and a spike timing-dependent plasticity network [8], implemented the speech control for robot, and carried out a physical experiment on the robot named Khepera. The results proved OCR model's effectiveness in speech control. Ruan, from Beijing University of Technology, built a computational model based on OCR theory using probabilistic automaton [9], and performed simulation of the Skinner's pigeon experiment, and the proposed model has better bionic learning ability. Then, his research team built an OCR algorithm using BP neural networks, and did the some simulation experiment on a two-wheeled self-balancing robot [10]. In the aspect of neurophysiology, O'Doherty *et al.* [11] designed a kind of OCR model, and through the fMRI technology, they found that ventral striatum in the basal ganglia affects reward and motivation, and dorsal striatum affects behavior and cognitive control.

Considering the basal ganglia neurophysiological basis of the OCR learning process, a kind of basal ganglia cognitive behavioral model based on OCR (OCR-based BCM) was established. In this model, the simulated annealing method was introduced to the winner-to-all mechanism of basal ganglia, and using the skinner's pigeon experiment, we successfully simulated the animal's gradual learning and adaptive performance.

2. OPERANT CONDITIONING REFLEX

2.1. Principle of Operant Conditioning Reflex (OCR)

Skinner's OCR theory (also called as instrumental learning or operant learning) is a kind of behavior changing

*Address correspondence to this author at the Dagunan Road, Tianjin, China; Postcard: 300222; Tel: 022-88181055; E-mail: c_j_223@163.com

process caused by stimulus. OCR is different from classics conditioning reflex (CCR). The object of OCR is originally the behavior of one's own accord, while CCR will make some individuals generate behavior from nonspontaneous reaction. OCR theory indicates that when an action makes the system develop in a better direction, or the action is right; then in the following process, the action's probability in the next same state will be increased, *i.e.* $p(a\,|\,s) = p(a\,|\,s) + \Delta p$. Otherwise, when an action makes the system develop in a worse direction, or the action is bad; then in the following process, the action's probability in the next same state will be decreased, *i.e.* $p(a\,|\,s) = p(a\,|\,s) - \Delta p$. After a period of operant condition training, the agent will get adapted to the operation environment. As per Björn Bremb's opinion on OCR learning, the prediction is very important in the whole instrumental learning process.

For a system that has not yet been fully understood, we can learn the past experience to predict its future behavior. An important advantage of prediction learning is that its training samples directly come from the time series of the real-time input signal without the need for special teachers. Based on this idea, we first established the following OCR learning model.

2.2. OCR Learning Model

Definition 1: A OCR learning model (OCRLM) can be expressed as a 8-tuple computational model OCRLM $= \langle S, A, f, \varphi, r, V(S,A), P, L \rangle$, and the meaning of each elements are as follows.

(1) S : OCRLM internal states sets, with $S = \{s_i \big| i = 0,1,2,\cdots,n\}$, S is a finite nonempty set including all possible states, s_i denotes the *i-th* state, and *n* is the number of states;

(2) A : OCRLM selectable action sets, with $A = \{a_i \big| i = 0,1,2,\cdots,m\}$, a_i denotes the *i-th* operant action, *m* is the number of selectable actions;

(3) f : OCRLM states transition function $f : S(t) \times a(t) \to S(t+1)$, *i.e.* state at t+1: $S(t+1) \in S$ is decided by $S(t) \in S$ and action $a(t) \in A$, generally decided by environment or system model;

(4) φ : OCRLM tropism mechanism $\varphi(t) = \varphi(S(t))$ denotes the system's orientation property at time t. As the definition of system states, the tropism is defined from biological significance. Environment determines the direction of biological evolution. The greater the tropism value is, the better it is. Different tropism functions can be defined according to the different situations;

(5) $r : r(t) = r[S(t), U(t)]$ is the reward after action $A(t)$. Based on the definition of tropism function, if $\varphi(t+1) > \varphi(t)$, indicates that the system is developing in the better direction, then $r = 1$; and if $j\ (t+1) < j\ (t)$, indicates that is developing in the worse direction, then $r = 0$;

(6) $V(S,A)$: OCRLM predication function, on one state S , $V(S,A) = \{v_i(S,a_i) \big| i = 0,1,2,\cdots,m\}$, can be seen as the action estimation to future reward discount sum;

(7) P : OCRLM probability vector from condition state to selective action, $P = [\,p(a_1), p(a_2), \cdots, p(a_m) \big| S] = [\,p_{a_1,S}, p_{a_2,S}, \cdots, p_{a_m,S}]$, and the action selection obeys the probability distribution of G $p_{a_j,S} = p(a = a_j\,|\,S) = \dfrac{e^{\beta V(S,a_j)}}{\sum_{a \in A} e^{\beta V(S,a)}}$, the meaning is that in the condition of *S*, OCRM operate behavior a_j ŒA as probability $p(a_j) \in \Gamma$, and $0 < p(a_m \big| s) < 1$, $\sum_{i=1}^{m} p(a_i \big| s) = 1$;

(8) L : OCRLM learning mechanism, $L : P(t) \to P(t+1)$. The updating is mainly completed by the updating of evaluation mechanism, and here it is implemented through updating the weights in the prediction function $V(S,A)$ by TD learning algorithm.

As we know, entropy in information theory is a physical quantity representing the degree of system uncertainties. And the bigger the information entropy, the greater the uncertainty of the system is. On one state, the entropy E_k is defined as follows:

$$E_k = -\sum_{i \in U(k)} \pi_k(i) \log \pi_k(i) \tag{1}$$

Where, $\pi_k(i)$ denotes the probability of selecting action *i* at the state *k*. Based on the entropy definition, we can define the condition entropy as follows:

Definition 2: Condition entropy $H(s_i)$ denotes the operation entropy of OCRLM on state $s_i \in S$:

$$H(s_i) = -\sum_{k=1}^{m} p(a_k \big| s_i) \log_2 (p(a_k \big| s_i)) \tag{2}$$

The basic principles of the entire OCR learning process can be briefly as follows: at time t , suppose the system state is $S(t) = s_i \in S$. According to the initial predication function and probability P vector to determine the selection probability of each operation behavior, and according to the competition mechanism on the basis of probability to choose operation behavior $a_k \in A$, carry out the action, and then the state is transformed into $S(t+1) \in S$. According to the tropism information, obtain the real-time evaluation information of this operation; and according to TD learning to update the predication network $V(S,A)$, find new prediction estimates and get a new probability vector. Continue the behavioral choice in the next round, until the updated network

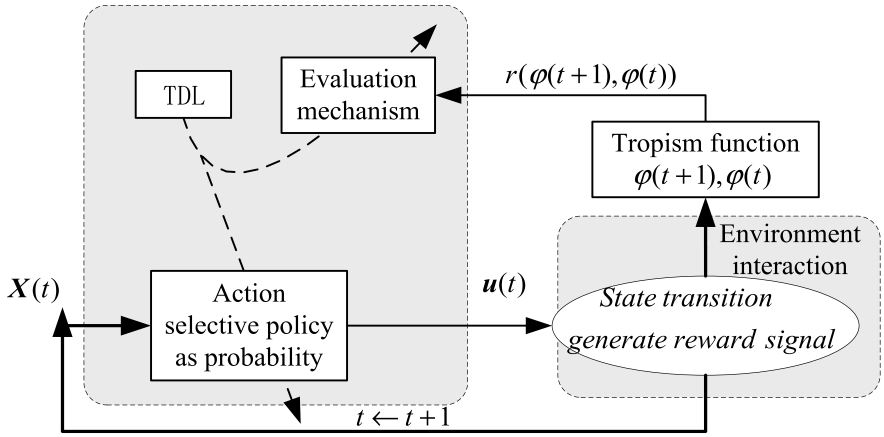

Fig. (1). The learning control mechanism based on OCR scheme.

can learn the most optimal operation. So, using OCR forms, the study process is over. The whole process can be represented in Fig. (**1**).

3. BASAL GANGLIA COGNITIVE BEHAVIORAL MODEL BASED ON OPEARNT CONDITIONING REFLEX THEORY(OCR-BASED BCM)

3.1. Basal Ganglia Cognitive Behavioral Model Design

Basal ganglia (BG) are composed of a series of nerve nuclei in deep brain. It is connected to the cerebral cortex, thalamus and brainstem. According to the anatomy and physiology, the BG's main function is to control voluntary movement, and it can perform behavior choice and reflex learning. Basal ganglia includes caudate and putamen (striatum), globus pallidus, substantia nigra and thalamic nuclei. The striatum can be divided into two complementary and phenotypically distinct compartments: the striosome and the matrisome according to different cholinesterase. Striatum is the structure for Basal ganglia receiving excitability input from the cerebral cortex (CC), while striosome only receives prefrontal excitatory afferent fibers project. Matrisome sends inhibitory efferent fibers to pallidum medial segment and the substantia nigra pars reticulate (SNR) forms the structure of the output of the basal ganglia. Striosome output to the Substantia Nigra Compacta, control and adjust the Substantia Nigra and Striatum pathways. Eventually CC-BG-thalamus-CC loop is formed. Among them, the action evaluation is done in striosome, action selection is done in matrisome. Dopamine signal from Substantia Nigra Compacta is used as the guidance of motion evaluation to improve the prediction to future reward. The aim is to get the better action.

Based on the OCR principle and BG neurophysiology, this paper established a cognitive behavioral model of the basal ganglia, which mainly includes: sensory cortex (SC), motor cortex (MC), Striatum(striosome, matrix)(STR), Substantia Nigra (SN), which can be expressed by formula (3).

$$\mathrm{BG} \sim (SC, MC, STR_{striosome}(CC, CC\text{-}STR_{synapsis}), STR_{matrix}, SN_{DA}) \tag{3}$$

Each abbreviated element has the following meaning:

SC	–	System state information sensed by sensory cortex
MC	–	Action information sensed by motor cortex
CC	–	State and action information sensed by cerebral cortex
$CC\text{-}STR_{synapsis}$	–	Cortico-striatal synaptic connections between the cerebral cortex and striatum
$STR_{striosome}$	–	Striosome output of the prediction of future reward discount sum
SN_{DA}	–	Dopamine from SN

Basal ganglia cooperate with cerebral cortex and thalamus and form the OCR learning mechanism. The whole learning flow is shown in Fig. (**2**). The solid line denotes the signal flow, while the dashed line denotes synaptic modification. In the scheme, thalamus implements two functions: one is generating the reward information and the other is transferring information.

3.2. STR Implementation Based on Recurrent Neural Network

3.2.1. Elman Network

Elman network is a kind of regression neural network, as shown in Fig. (**3**). From the view point of system, it not only contains the input layer, hidden layer and output layer nodes, but also contains the feedback node with the same number of hidden layer nodes, earlier used to record hidden layer unit output value of the moment, which can be considered as a delay operator. The signal can be passed back and forth between neurons, when the entire network is in a constantly changing dynamic condition, which can more directly and vividly reflect the system's dynamic characteristic in the process of calculation. This network has better dynamic behavior and computing power than the forward neural network. Regression network is used, which can enhance the ability of dealing with dynamic information, to make it more suitable for the stability control of complex systems. This kind of learning pattern is similar to human brain, in which the memory of the new information will not affect the information already contained; thus it can reflect the stability of the human brain memory.

Network input vector $IN \in \mathbf{R}^{r\times 1}$ denotes the state-action pair at time t, output is $Y \in \mathbf{R}^{m\times 1}$, the node number of hidden layer is h, input vector of hidden layer is $\mathbf{O}(t) = \left[o_1(t), o_2(t), \cdots, o_h(t)\right]^{\mathrm{T}} \in \mathbf{R}^{h\times 1}$, output vector of hidden layer is $\mathbf{H}(t) = \left[h_1(t), h_2(t), \cdots, h_h(t)\right]^{\mathrm{T}} \in \mathbf{R}^{h\times 1}$, the network connection weights are $W^{(1)} \in \mathbf{R}^{h\times r}, W^{(2)} \in \mathbf{R}^{h\times h}, W^{(3)} \in \mathbf{R}^{m\times h}$ respectively, output layer activation function is the linear weighted function, then we can get the following formula:

$$Y(t) = W^{(3)}\mathbf{H}(t) \tag{4}$$

$$h_j(t) = \sigma(o_j(t)) = \frac{1}{1+\exp(-o_j(t))} (j = 1, 2, \cdots, h) \tag{5}$$

$$\mathbf{O}(t) = (W^{(2)})^{\mathrm{T}}\mathbf{H}(t-1) + (W^{(1)})^{\mathrm{T}} IN(t) \tag{6}$$

$\sigma(\cdot)$ is the nonlinear activation function of hidden layer, here $\frac{d\sigma(x)}{dx} = \sigma(x)[1-\sigma(x)]$ is Sigmoid function. Through the formulas (4-6), we can get that the output of hidden layer can be seen as the state of nonlinear dynamic model. This kind of regression neural network can acquire the dynamic characteristic of nonlinear system.

3.2.2. STR Implementation

The cortex information includes sensory cortex information and motor cortex information, so

$$CC = [SC; MC] \tag{7}$$

$$STR_{striosome} = STR(CC, CC\text{-}STR_{synapsis}) \tag{8}$$

Where $STR_{striosome}$ outputs the prediction information of future reward discount sum.

$$STR_{striosome}(t) = r(t+1) + \gamma r(t+2) + \gamma^2 r(t+3) + \cdots \tag{9}$$

So, the future reward discount sum at time $t-1$ is:

$$STR_{striosome}(t+1) = r(t+2) + \gamma r(t+3) + \gamma^2 r(t+4) + \cdots \tag{10}$$

Striatum output is implemented by Elman network above, where the hidden layer represents the granulosa cells in the cerebral cortex and the neural network weight represents the cortico-striatal synaptic connection. By formulas (9) and (10), we can get $STR_{striosome}(t) = r(t+1) + \gamma STR_{striosome}(t+1)$, which shows that the action evaluation value $STR_{striosome}(t)$ at time t can be denoted by $STR_{striosome}(t+1)$. But in the beginning of prediction some error must exist, and the estimated evaluation value $STR_{striosome}(t)$ expressed by $STR_{striosome}(t+1)$ is not equal to the actual $STR_{striosome}(t)$, and thus the substantia nigra generates the dopamine response, denoted by SN_{DA} as follows:

$$SN_{DA} = r(t+1) + \gamma STR_{striosome}(t+1) - STR_{striosome}(t) \tag{11}$$

3.3. Cortico-Striatal Synaptic Modification

The dopamine is produced by substantia nigra. Its feedback to striatum is used to modify the cortico-striatal synaptic and Nigra-Strio loop form. The modification mechanism is shown through the formulas (12) and (13).

$$CC\text{-}STR_{synapsis}(t) \leftarrow CC\text{-}STR_{synapsis}(t) + \Delta CC\text{-}STR_{synapsis}(t) \tag{12}$$

$$\Delta CC\text{-}STR_{synapsis}(t) = \alpha \cdot SN_{DA} \cdot \frac{\partial STR}{\partial CC\text{-}STR_{synapsis}} \tag{13}$$

3.4. Action Selection Mechanism in Matrix

In the operant learning process of basal ganglia cognitive behavior model, the matrix in striatum stimulates mechanism of action selection. During the operant learning, selecting action in probability is very important. Here, we use the Boltzmann Gibbs probability distribution to define the behavior choice probability. The expression is given through formula (14).

$$P(a = a_i \mid SC(t)) = \frac{e^{STR_{striosome}(SC(t), a_k)/T}}{\sum_{a\in A} e^{STR_{striosome}(SC(t), a)/T}} \tag{14}$$

Where, T>0 is the temperature constant. The larger the T is, the greater the degree of random behavior choice. When T is near to zero, the probability of choosing the action with the most value $STR_{striosome}(SC(t), a_k)$ is near to one. T is decreased with time, which denotes that during the learning process, the system accumulates more and more knowledge. The system gradually evolves into a deterministic system from an uncertainty system. In the beginning, in order to express the activity of biological systems, the T should not be set too small.

$$\begin{cases} T_0 = T_{\max}, \\ T_{t+1} = T_{\min} + \beta(T_t - T_{\min}) \end{cases} \tag{15}$$

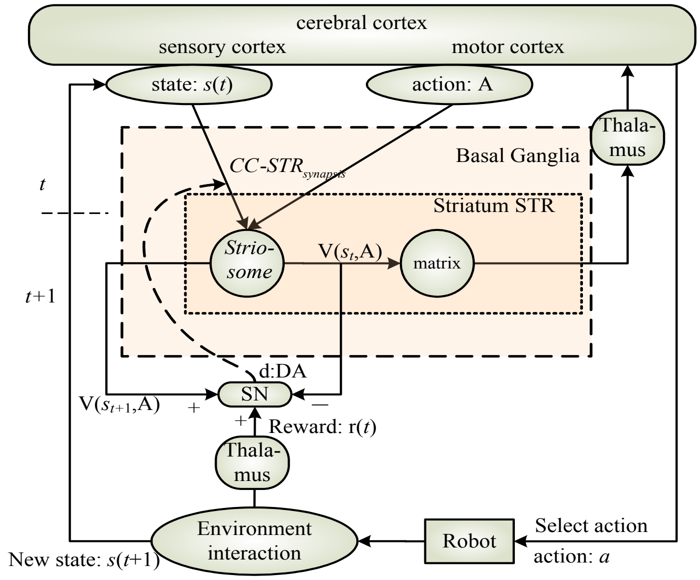

Fig. (2). The learning control mechanism based on OCR scheme.

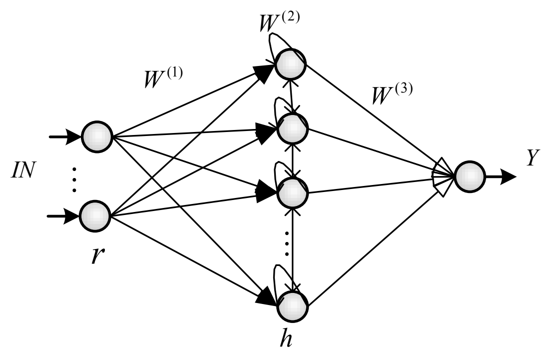

Fig. (3). Elman network topology.

Where $0 \le \beta \le 1$ is the annealing factor.

3.5. OCR Learning Process in Basal Ganglia

Step 1: Initialization $t = 0$, $CC\text{-}STR_{synapsis}(0) = 0$, so $\forall s \in S, p(a_k \mid s) = \frac{1}{m}(k = 1,2,\cdots,m)$, where m is the action number. This means that in the early state, the basal ganglia do not have any predetermined decisions. Input the beginning $SC(0)$, actions set A and the learning coefficient of OCR learning mechanism.

Step 2: Get the output of Striatum and the selected action. Compute each $STR_{striosome}(SC(t), a_k)$ for all the actions, and get the probability for selecting each action based on formula (14). Output action $a(t)$ through the action selection mechanism in the matrix of striatum and store $STR_{striosome}(SC(t), a(t))$.

Step 3: Carry out the selected action $a(t)$, the state is transferred to $SC(t+1)$, get the immediate reward r_{t+1}. Then, select $a(t+1)$ as Step 2, and get striatum predication value $STR_{striosome}(SC(t+1), a(t+1))$ at time $t+1$.

Step 4: Form OCR. Based on formulas (12) and (13), the cortico-striatal synapsis is modified. By decreasing the temperature based on formula (15), and changing the action selection probability, the OCR is formed.

Step 5: Time is updated.

4. APPLIACTION OF OCR-BASED BASAL GANGLIA COGNITIVE BEHAVIORAL MODEL IN PIGEON EXPERIMENT

In order to verify the OCR behavior property of the proposed cognitive model, we applied the model in the Skinner pigeon experiment.

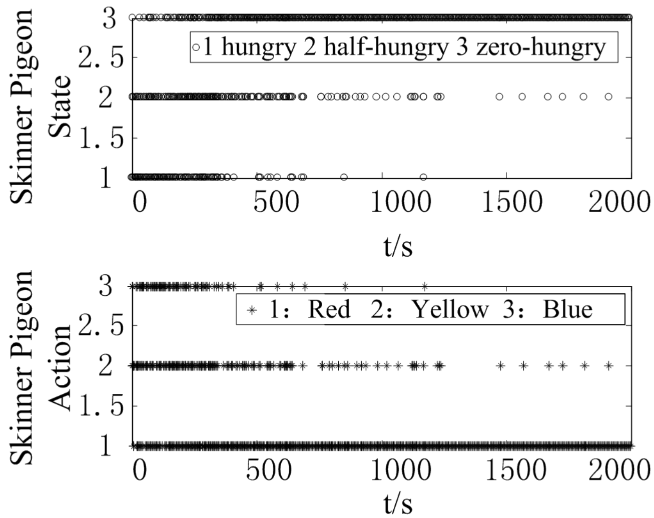

Fig. (4). Pigeon state and action results.

4.1. Skinner Pigeon Experiment Model Establishment

Skinner pigeon experiment is a classic experiment about OCR, which was proposed by Skinner. The experiment details are as follows. A pigeon is put in a box with three buttons. Initially, the pigeon can peck the buttons randomly, but when it peck red button, it will get some food (a kind of positive reinforcement stimulus). When it peck yellow button, there is no any stimulus, and when it peck blue button, the pigeon will get electric shock (a kind of negative reinforcement stimulus). After a period of time, the frequency of pecking red button is apparently higher than other two buttons. The pigeon get some knowledge through active action, and it can autonomously get food when it is hungry and avoid electrical shock.

In order to build Skinner pigeon experiment mathematical model, we firstly code the state and action.

4.1.1. State and Action Coding

Supposing the pigeon have three states: hungry ($s_0 = 1$), half-hungry($s_1 = 2$) and zero-hungry($s_2 = 3$). The pigeon is in a good state when the state value is big. Pigeon have three actions: pecking the red button, yellow button and blue button, separately coded as $a_0 = 1, a_1 = 2, a_2 = 3$ respectively.

4.1.2. State Transition

The pigeon experiment's state transition equation f : $S(t) \times A(t) \rightarrow S(t+1)$ can be defined as follows:

$$\begin{aligned} &f(s_0,a_0)=s_1 \quad f(s_1,a_0)=s_2 \quad f(s_2,a_0)=s_2 \\ &f(s_0,a_1)=s_0 \quad f(s_1,a_1)=s_0 \quad f(s_2,a_1)=s_1 \\ &f(s_0,a_2)=s_0 \quad f(s_1,a_2)=s_0 \quad f(s_2,a_2)=s_0 \end{aligned} \tag{16}$$

4.1.3. Tropism Function and Reward Mechanism

We defined the tropism function as formula (17) for Skinner pigeon.

$$\varphi(t) = s(t) \tag{17}$$

The tropism function shows that when the pigeon is in zero-hungry state, $s(t) = 3, \varphi(t) = 3$, it has a better tropism state. When the pigeon is in hungry state, $s(t) = 1, \varphi(t) = 1$, the tropism is worst.

Based on tropism function, we can define the reward mechanism. If $\varphi(t+1) \geq \varphi(t)$, it denotes that the system is changing to a better state, $r = 1$, and if $\varphi(t+1) < \varphi(t)$, it denotes that the system is changing to a bad state, $r = 0$, *i.e.*

$$r_{t+1} = \begin{cases} 1(penalty) & if(\varphi(t+1) \geq \varphi(t)), \\ 0(reward) & else. \end{cases} \tag{18}$$

4.2. Simulation Experiment and Analysis

Set the parameter during simulation: $r = 2, h = 5, m = 1$, Maxstep=2000, sample time $T_c = 1s$. Simulation experiment is done in MATLAB environment. During the process, we record the state and selected action of pigeon at each moment, as shown in Fig. (**4**).

A statistical analysis of behavior of selecting three buttons is done based on six statistical results, respectively, at the time of: 200s / 400s / 600s / 800s / 1200s / 2000s. The six statistical results are shown in Fig. (**5**). Fig. (**6**) shows the probability changes of pigeon pecking of different colored buttons.

In Figs. (**4**-**6**), it is shown that in the beginning of the training phase, the initial probability of selecting different operating behavior is the same, so the number of pigeons pecking three buttons is basically the same, which is equal to 0.3333. But along with the random environment interactions,

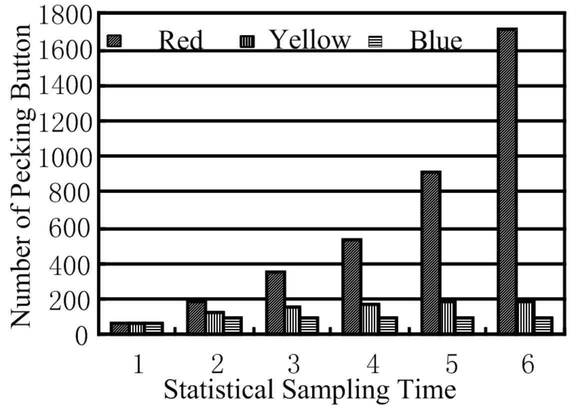

Fig. (5). Number of selected three action.

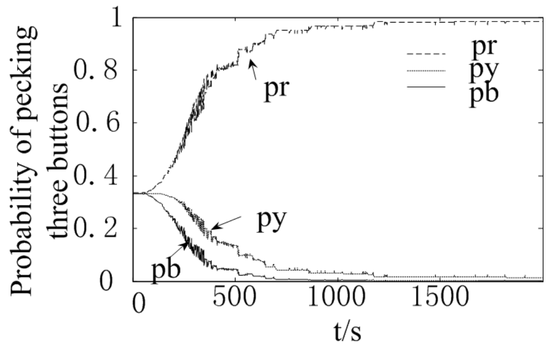

Fig. (6). The change of probability.

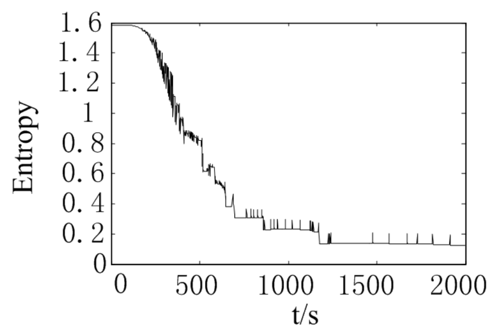

Fig. (7). The change of entropy.

the cortico-striatal synaptic weights were changed unceasingly, causing the change of behavior selection probability. The number of red button selection probability for pigeon is gradually increased, and the probability number of pecking yellow and blue button is gradually reduced. When in the 2000s, probability of pecking red button is far greater than other two buttons, thus the OCR is formed in the basal ganglia.

By formula (2), system entropy can be calculated at each moment, and the entropy change with time is shown in Fig. (**7**). Through the changing trend of entropy, it can be seen that at the initial stage, randomness of pigeon behavior choice is bigger. Over a period of operating conditions training of basal ganglia, neural structures (mainly referring to the cortico-striatal synaptic weights) change, and the entropy is gradually reduced. It also illustrates that learning is a process marked by decreased entropy, which is an orderly self-organization process.

CONCLUSION

Based on the OCR principle, a basal ganglia cognitive behavioral model has been proposed. Firstly, by applying prediction mechanism, an OCR learning model is established, which corresponds with the behavior selection mechanism of basal ganglia. And then, the basal ganglia model based on OCR is proposed. In the proposed model, a cortico-striatal synapsis model is implemented using Elman neural network weights. Input information from cortex and sensory which is a state-action pair. Striosome in striatum outputs the prediction of actions, and matrix in striatum selects action. The action is delivered to motor cortex and puts into the environment, then the cortico-striatal-thalamic-cortical (CSTC) circuit is formed. Cortico-striatal synapsis model is modified by the dopamine from substantia nigra, and hence the substantia nigra-striatal circuit is formed.

The proposed basal ganglia cognitive behavioral model based on OCR is suitable for the limited cognitive behavior learning, and there is no limit to the number of system status. By applying the bionic learning model, Skinner pigeon experiment is simulated in MATLAB. Through simulation, it can be seen that the constructed model of basal ganglia effectively simulates the biological phenomenon of OCR, and shows better self-organizing, adaptive and self-learning ability, which provides reference for the further research on cognitive model. In the following research, this proposed model will be used in robot to further verify the effectiveness of the model.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

We acknowledge support from National Natural Science Foundation of China (No. 61403282), Tianjin City High School Science & Technology Fund Planning Project (No.20130807), Tianjin University of Technology and Education Project (No. KJY1311, No.KYQD13004). We are thankful to the reviewers for helpful comments to improve the quality of this paper.

REFERENCES

[1] B.E. Rosen, J.M. Goodwin, and J.J. Vidal, "Machine operant conditioning," In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1988*, vol. 3, pp. 1500-1501, 1988.

[2] E. Zalama, P. Gaudiano, J.L. Coronado, "Obstacle avoidance by means of an operant conditioning model," In: *International Workshop on Artificial Neural Networks Malaga-Torremolinos*, Spain, vol. 930, pp. 471-477, 1995.

[3] P. Gaudiano, C. Chang, "Adaptive obstacle avoidance with a neural network for operant conditioning: Experiments with real robots," In: *Proceedings of 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation, CIRA'97*, pp.13-18, 1997.

[4] B. Brembs, "brembs.net: Research on Learning, Memory and Evolution," Available at: http://brembs.net/.2014.

[5] B. Brembs, "Spontaneous decisions and operant conditioning in fruit flies," *Behavioural Processes*, vol. 87, no. 1, pp. 157-164, 2011.

[6] E. Zalama, J. Gomez, M. Paul, and J.R. Peran, "Adaptive behavior navigation of a mobile robot," In: *IEEE Transactions on Systems, Man, and Cybernetics, Part A-Systems and Humans*, vol. 32, no. 1, pp.160-169, 2002.

[7] K. Itoh, H. Miwa, M. Matsumoto, M. Zecca, H. Takanobu, S. Roccella, M. C. Carrozza, P. Dario, and A. Takanishi, "Behavior model of humanoid robots based on operant conditioning," In: *5th IEEE-RAS International Conference on Humanoid Robots, 2005*, pp. 220-225, 2005.

[8] T. Tadahiro, and S. Tetsuo, "Incremental acquisition of behaviors and signs based on a reinforcement learning schemata model and a spike timing-dependent plasticity network," *Advanced Robotics*, vol. 21, no. 10, pp. 1177-1199, 2007.

[9] X.G. Ruan, J.X. Cai, and L.Z. Dai, "Compute model of operant conditioning based on probabilistic automata," *Beijing Gongye Daxue Xuebao/Journal of Beijing University of Technology*, vol. 36, no. 8, pp. 1025-1030, 2010.

[10] H. Ren, and X. Ruan, "Self-balance control of two-wheeled robot based on Skinner's operant conditioned reflex," *Control Theory & Applications,* vol. 27, no. 10, pp. 1423-1428, 2010.

Received: September 16, 2014 Revised: December 23, 2014 Accepted: December 31, 2014