

A Novel Curiosity-Driven Perception-Action Cognitive Model

Jing Chen^{1,*}, Bing Li¹ and Li Li¹

¹ School of Information Technology Engineering, Tianjin University of Technology and Education, 300222, Tianjin, China

*Corresponding author

Abstract—Aiming at agents' autonomous cognitive problems in unknown environment, a novel curiosity-driven perception-action cognitive model is proposed, which simulates intrinsic motivation cognitive mechanism based on the curiosity in psychology, and cognitive processes from perception to action is realized by probabilistic action selection mechanism. Information entropy illustrates that it can achieve better cognition to action applying the proposed model, and it indeed reflects the biological cognitive processes. Comparing simulation results using SPE shows that this method is effective.

Keywords—perception-action; cognitive model; skinner pigeon; curiosity

I. INTRODUCTION

With the study of cognitive robots, "cognitive control" has gradually replaced the "intelligent control"[1], which is mainly used to solve the control problem by simulating the biological cognition mechanism. According to the Piaget's cognitive development theory, sensorimotor is the primary stage of cognitive development. Human's sensorimotor (also can be called Perception-Action) function is the result of cognitive development, which is gradually formed, developed and perfected. In 2013, Wu Xuan and Xiaogang Ruan present a psychology model called skinner automata, which implemented the theory of operant conditioning reflex(OCR), and is used in equilibrium learning of self-balancing robot, and showed the psychology model effectiveness in the incremental learning process[2,3]. In 2015, Huang Jing put forward a kind of artificial sensorimotor system with OCR function, and realized the cognitive process of the mapping relationship between the state and action, and made a comparison verify using two experiment from psychology and cybernetics field[4].

In 2013, Cutsuridis et al proposed a cognitive control architecture for the perception-action cycle in robots and agents, which is composed of a large number of neural computing mechanism, and this view is strongly supported by the evidence of brain experimental research[5]. In the intrinsic motivation research framework, in 2010, Baranes et al. proposed the adaptive target generation algorithm based on curiosity as the mechanism of intrinsic motivation, which is used in the autonomous learning of robot[6]. In 2012, action-perception loop based on internal model is used in the research of exploration mechanism in the process of cognition of agents without the external reward feedback[7]. Baranes and Oudeyer proposed an adaptive target generation - robust intelligent

adaptive curiosity framework as the intrinsic motivation of the target exploration mechanism[8].

Self-learning is a main motivation of the behavior making, and the exploration strategy in the intrinsic motivation plays an important role in the process of cognition. As one of the main factors that increase the intrinsic motivation, curiosity has played an important role in the learning process of "perception-action" loop. Curiosity mainly includes sensory curiosity(SC) and cognitive curiosity(CC). A new model named as curiosity-driven perception-action cognitive model (CPACM) is proposed in this paper, which apply the sensory curiosity (curiosity cognitive) in the learning of the perception-action loop. Using the classical skinner pigeon experiments, the effectiveness is verified in the simulation environment of MATLAB.

II. CURIOSITY-DRIVEN PERCEPTION-ACTION COGNITIVE MODEL

A. Structure Design

On the basis of operant conditioning cognitive models with probabilistic behavior making, in the formation process of "perception-action" mapping, the proposed model increase the curiosity-driven module, and the cognitive structure is shown in Figure I.

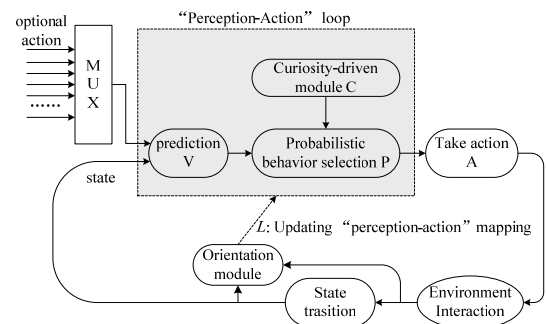


FIGURE I. COGNITIVE STRUCTURE OF CPACM

B. Mathematical Description

The proposed curiosity-driven perception-action cognitive model can be described as a 9-tuple computational model $CPACM = \langle S, A, f, \phi, r, V(S, A, t), C(S, t), P(t), L \rangle$. Every elements are explained as follows.

S is the internal state set of CPACM, $S = \{s_i | i = 1, 2, L, n\}$;

A is the optional action set of CPACM $A = \{a_i | i = 1, 2, L, m\}$;

f is the state transition of CPACM, $f: S(t) \times a(t) \rightarrow S(t+1)$, which is almost determined by environment or system model;

φ is the orientation mechanism, $\varphi(t) = \varphi(S(t))$, which denotes the state orientation at time t , and can be defined based on the specific case, and the bigger of the value, the better of the system;

$r: r(t) = r[S(t), A(t)]$ is the reward from state $S(t)$ to $S(t+1)$ after making the action $A(t)$ at time t ;

$V(S, A, t)$ is the prediction of CPACM, $V(S, A, t) = [v_1, v_2, L, v_m] = [v(S, a_1, t), v(S, a_2, t), L, v(S, a_m, t)]$ is a vector;

$C(S, t)$ is the curiosity function, which denotes the state curiosity at time t . It is a monotone decreasing function with respect to time t . In the condition of the same state, curiosity is declined over time, in line with the characteristics of biological learning;

$P(t)$ is the probability vector from the condition of the state to optional behavior for the CPACM, $P(t) = [p(a_1, t), p(a_2, t), L, p(a_m, t) | S] = [p_{a_1, S}(t), p_{a_2, S}(t), L, p_{a_m, S}(t)]$, action selection probability

$p_{a_j, S}(t) = p(a = a_j | S, t) = e^{V(S, a_j, t)/C(S, t)} / \sum_{a \in A} e^{V(S, a, t)/C(S, t)}$ denotes that agent choose the action $a_j \in A$ as $p(a_j, t) \in P$ in the condition of state S at time t , and $0 \leq p(a_m | s, t) \leq 1$,

$$\sum_{i=1}^m p(a_i | s, t) = 1;$$

L denotes the updating of "Perception-Action" mapping of CPACM, $L: PA(t) \rightarrow PA(t+1)$. This updating mechanism is implemented through changing the prediction critic network, which use the TD(0) method to update the weight of $V(S, A, t)$ of CPACM, that is $\Delta W(t) = \alpha \cdot \delta_{TD(0)} \cdot \partial V(IN, W) / \partial W$, where, $\delta_{TD(0)} = r(t+1) + \gamma V(t+1) - V(t)$, and then the action selection probability is changed.

The cognition process could be concluded as follows. At time t , the state of agent is $S(t) = s_i \in S$, and based on the initial prediction critic function $V(S, a_i, 0)$, curiosity parameter and P vector, the behavior selection probability of each action could be decided, and then select the action $a_k \in A$ as probability to act on the environment, and state is transformed to $S(t+1) \in S$, and then get the immediate critic information r , and then $V(S, A, t)$ is updated based on TD(0) algorithm to formulate new prediction estimation value, and update the curiosity value at this state, and the get new probability

vector P , and the first loop of cognition is complemented. Looping as the above process, the "perception-action" loop based on curiosity-driven intrinsic motivation is formed, and the agent learns the behavior selection mechanism, so the cognition process is over.

In order to characterize the certainty degree of the system, we use information entropy to measure, and conditional entropy is used here to describe the behavior information entropy in a defined state (in bit), which is defined as follows.

$$H(S, t) = - \sum_{i=1}^m p_{a_i, S}(t) \log_2 p_{a_i, S}(t) \quad (1)$$

where, $p_{a_i, S}(t)$ represents probability of selection behavior a_i in the condition of state S at time t and satisfies $\sum_{i=1}^m p_{a_i, S}(t) = 1$, which denotes that the bigger of the information entropy, the higher of the system's uncertainty degree.

The proposed cognitive model express the process from random to certainty, and reflect the gradual learning characteristics similar to biology.

III. EXPERIMENTAL SIMULATION AND ANALYSIS

In order to verify the proposed curiosity-driven "perception-action" cognitive model's behavioral learning ability, we make a simulation research using typical experiment (skinner pigeon experiment) in behavioral learning theory, and make a experimental comparative analysis with the operant conditioning models to illustrate the effectiveness of the proposed method.

A. Skinner Pigeon Experiment

Skinner Pigeon Experiment (SPE), which has been described in literature [3] in detail is a classic experiment in behavioral learning theory. In this experiment, he pigeon is placed in a designed box named skinner box. In the face of three different colored buttons, that is red, yellow and blue color, the pigeon makes an action selection of pecking at the three buttons. When it pecks at different button, it will get different response. If pecking at red button, pigeon will get food for reward, and pecking at yellow button does not respond, and pecking at blue button, pigeon will get electric shock as punishment. At the beginning, number of pigeon pecking the three buttons is almost equal. But after a period of cognitive, the number of pigeon pecking at the red button is significantly increased, and the pigeon learned the behavior of pecking at the red button autonomously to get food reward this behavior.

In order to facilitate simulation of the experiment, the state and behavior was coded, and the simplified discrete mathematical model is established. The state transition diagram is shown in Figure II, in which the circle denote the state in which and arrows indicate the transition between two states, and the value above the arrow indicates the action taken.

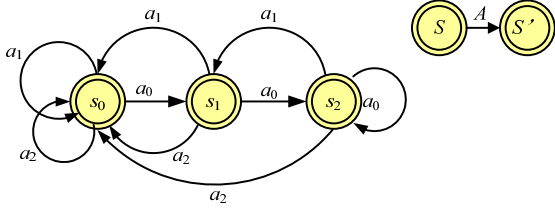


FIGURE II. STATE TRANSITION DIAGRAM OF SPE

In Figure 2, $s_0 = 1$ is the state of pain; $s_1 = 2$ is the state of desire to obtain food; $s_2 = 3$ is the state of satisfied. The bigger the state value, the better the pigeon's state. The orientation of the pigeon is satisfying itself need in the maximum extend and the better the state, the bigger the orientation value. The three action of pigeon is $a_0 = 1$, $a_1 = 2$, $a_2 = 3$, which respectively denotes pecking at red, yellow and blue button.

To express prediction function V , a recurrent neural network is used, and the input vector is $IN \in \mathbf{R}^{r \times 1}$, output is $Y \in \mathbf{R}^{m \times 1}$, hidden nodes number is h , the input vector of hidden layer is $\mathbf{O}(t) = [o_1(t), o_2(t), L, o_h(t)]^T \in \mathbf{R}^{h \times 1}$, which could be seen as the excitation function of network internal state. The output vector of hidden layer is $\mathbf{H}(t) = [h_1(t), h_2(t), L, h_h(t)]^T \in \mathbf{R}^{h \times 1}$, which denotes the internal state set of the network. The weights of this recurrent neural network is $W^{(1)} \in \mathbf{R}^{h \times r}, W^{(2)} \in \mathbf{R}^{h \times h}, W^{(3)} \in \mathbf{R}^{m \times h}$, and the activation function of output layer is a linear weighting function as follows.

$$Y(t) = W^{(3)} \mathbf{H}(t) \quad (2)$$

The internal state transition function is $\mathbf{H}(t) = g(IN(t), \mathbf{H}(t-1), W^{(1)}, W^{(2)}) = [h_1(t), h_2(t), L, h_h(t)]^T$, and the excitation function of internal state is $\mathbf{O}(t) = (W^{(2)})^T \mathbf{H}(t-1) + (W^{(1)})^T IN(t)$, where, $h_j(t) = 1/(1 + \exp(-o_j(t)))$ ($j = 1, 2, L, h$). The perception state and action in SPE are all one dimension variable, so $IN = [x, a]^T \in \mathbf{R}^{2 \times 1}$.

In order to express state curiosity of pigeon, we make a statistics of the same state in the learning process, denoting as $Num(S, t)$, which is a monotone decreasing function with respect to time t , shown as formula (3).

$$C(S, t) = \frac{k_1}{1 + e^{k_2 \cdot Num(S, t)}} \quad (3)$$

where, $k_1 > 0, k_2 > 0$ is the adjustment factor of curiosity parameter, satisfy the following condition. When $Num(S, t) \rightarrow \infty$, $C(S, t) \rightarrow 0$, that meets the characteristic of biological cognition.

Define the reward mechanism based on the orientation as $\varphi(t) = S(t)$, and if $\varphi(t+1) \geq \varphi(t)$, $r = 0$. If $\varphi(t+1) < \varphi(t)$, $r = -1$.

For the SPE, the cognitive flow based on the proposed model is shown as Figure III, where, $step_{max}$ is the max run step during simulation. The initial state of pigeon is the state of desire to obtain food.

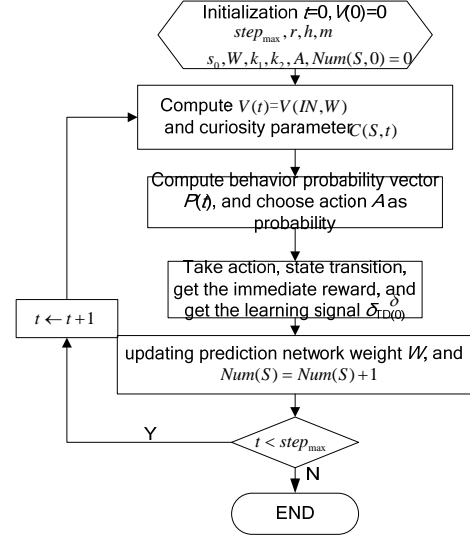
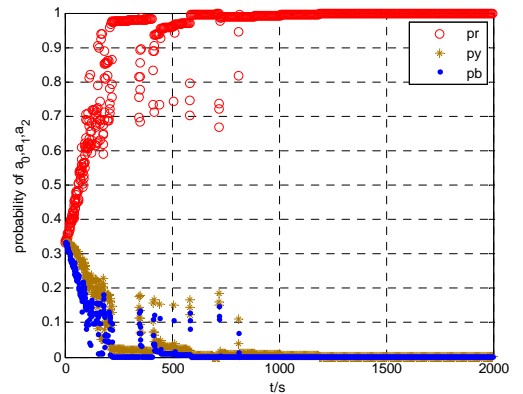


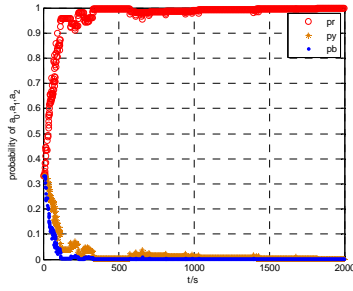
FIGURE III. COGNITIVE FLOW BASED ON THE PROPOSED CPACM MODEL.

B. Simulation Results and Analysis

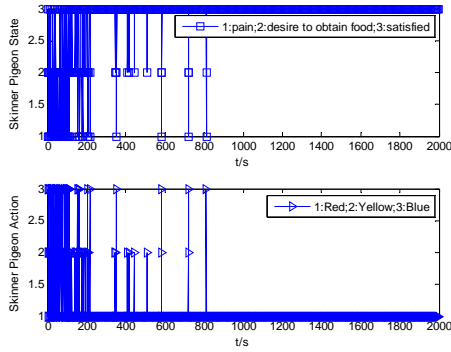
Using the CPACM cognitive models proposed in this paper and OCR model, cognitive process of SPE were simulated. The sampling time is 1s, and we record the data of the three action's selection probability, skinner pigeon's state and action taken at each second. Comparison of the results is shown in Figure IV.



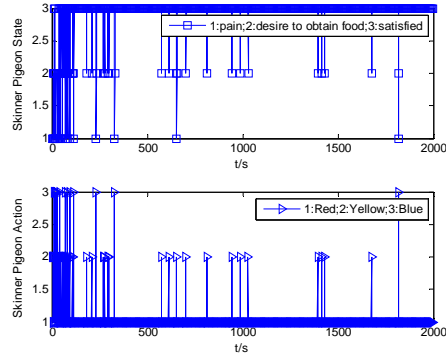
a) action selection probability changing using CPACM model with curiosity



b) action selection probability changing using OCRM model without curiosity



c) skinner pigeon's state and action taken changing using CPACM model with curiosity



d) skinner pigeon's state and action taken changing using OCRM model without curiosity

FIGURE IV. SKINNER PIGEON COGNITION RESULTS COMPARISON OF CPACM AND OCRM

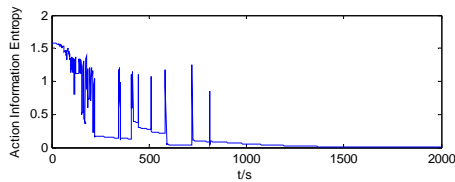


FIGURE V. BEHAVIOR INFORMATION ENTROPY CURVE OF SKINNER PIGEON.

As it can be seen from Figure IV a) and b), probability of skinner pigeon pecking at red button are gradually changing from 0.3333 to 1, ie pigeon selects the behavior of pecking at the red button with probability 1 eventually. Due to the

presence of curiosity in CPACM, in the initial moments pigeon has a certain curiosity at three states, so that a certain degree of behavior select probability fluctuation of skinner pigeon occur compared with OCRM method. By observing Figure IV c) and d) we will find that although the probability of CPACM method appears shock, but on the late stage skinner pigeon behavior and state have reached a more determined state. But in the sudden small probability events of OCRM method, behavioral choices certainty decreased. From Figure IV d) it would be observed that there is also the cases of choosing yellow or blue button, that is because in the early stages of learning, curiosity does not work, lack of cognition under certain states and behavior, cognition process is affected.

Figure V shows a variation of the system behavior information entropy. By entropy's trend tends to 0, we can conclude that the cognition of the system is a evolution process from uncertainty to certainty.

IV. CONCLUSION

In this paper, aiming at agents' autonomous cognitive problems in unknown environment, a novel curiosity-driven perception-action cognitive model is proposed, which simulates intrinsic motivation cognitive mechanism based on the curiosity in psychology, and cognitive processes from perception to action is realized by probabilistic action selection mechanism. In order to verify the proposed curiosity-driven "perception-Action" cognitive model's behavioral learning ability, we make a simulation research using typical SPE in behavioral learning theory, and make a experimental comparative analysis with the OCRM to illustrate the effectiveness of the proposed method. Information entropy illustrates that applying the proposed model it can achieve better cognition, and it indeed reflects the biological cognitive processes. Comparing simulation results show that this method is effective.

ACKNOWLEDGMENT

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), and thank for the reviewers' helpful comments to improve the quality of this paper.

REFERENCES

- [1] Bellas F, Caamaño P, Faiña A, et al. Dynamic learning in cognitive robotics through a procedural long term memory[J]. *Evolving Systems*, 2014, 5(1):49-63.
- [2] Wu X, Ruan X G, Zhang X, et al. Study of Skinner Automaton Implemented on a Two-Wheeled Robot[C]. *3rd International Conference on Electric and Electronics*, 2013: 253-256.
- [3] RUAN XiaoGang, WU Xuan. The skinner automaton: A psychological model formalizing the theory of operant conditioning[J]. *Science China(Technological Sciences)*, 2013, 11:2745-2761.
- [4] HUANG Jing, RUAN Xiao-gang, YU Nai-gong, ZHANG Xiao-ping, WEI Ruo-yan, FAN Qing-wu. Artificial sensorimotor system with operant conditioning function[J]. *Control theory and application*, 2015, 32(5):674-681.

- [5] Cutsuridis V, Taylor J G. A cognitive control architecture for the perception–action cycle in robots and agents[J]. *Cognitive Computation*, 5(3): 383-395, 2013.
- [6] Baranes A, Oudeyer P Y. Intrinsically motivated goal exploration for active motor learning in robots: A case study. 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 1766-1773, 2010.
- [7] Little, D. Y., Sommer, F. T. Learning and exploration in action-perception loops[J]. *Frontiers in neural circuits*, 7: 37-37, 2012.
- [8] Baranes, A., Oudeyer, P-Y. Active Learning of Inverse Models with Intrinsically Motivated Goal Exploration in Robots[J], *Robotics and Autonomous Systems*, 61(1):49-73, 2013.