2010 3rd International Conference on Computer and Electrical Engineering (ICCEE 2010)
IPCSIT vol. 53 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2012.V53.No.1.50

Bionic Robot Control based on Advanced Updating

Tao Shi[+] and Zhikun Chen

College of Computer & Automation Control
Hebei Polytechnic University
Tangshan, China

Abstract. In view of the balance control problem of self-balance robot, the reinforcement learning mechanism based on the advanced updating is proposed as a self-balance bionic learning algorithm of the robot. This algorithm can choose optimal behavior above the baseline with the certain probability by using baseline in the advanced updating and unifying the probability curiosity mechanism in reinforcement learning, so that the robot can obtain the bionic self-learning skills like creature under the unknown environment, and realize the bionic self-balance controls of the robot. Finally, the simulation experiment is made. The result indicates that the robot can obtain the stronger self-learning skills about the balance control and the better dynamic performance from this learning algorithm of the reinforcement learning based on the advanced updating and it manifests the bionic characteristic about the advanced updating, and gains the higher research and the application value.

Keywords: Advanced updating; reinforcement learning; curiosity mechanism; bionic; robot

1. Introduction

The reinforcement learning establishes a bridge between organism and machine, and the advanced updating just manifests the substantive characteristics of the reinforcement learning. The bionic learning algorithm of the reinforcement learning mechanism based on the advanced updating is adopted, through this algorithm, and some life characteristic of the creature can be expressed by the reinforcement learning model, and applied this characteristic model on the robot, so the robot may display some certain biology characteristics. Endow the robot with creature's intelligence, and realize the intellectualization of the robot, which is the key point of this article.

Reinforcement learning is a kind of mathematical description of the phenomenon conditioning reflex in the biology, they are extremely close contact. The conditioning reflex mostly researches choice behavior of creature, and this behavior is suitable for expressing the emotion and the environment. Through the alternation with environment, the robot can learn some kind of ability, which is the most important characteristics among the bionic learning. In recent years, the condition reflection already has been carried out massive research about the aspect in the robot cognition biological modeling. In 1995, Zalama and his colleague designed the arithmetic about avoiding barrier based on the condition reflection theory, and they have carried on the simulation on the computer and obtained good simulation result [1]; In 1997, Gaudiano and his colleagues of Boston University nerve robot laboratory applied the condition reflection slgorithm designed by Zalama et al. to the Pioneer 1 and the Khepera, and they carried on the avoiding barrier experiment and obtained good control result [2]. In 2005, Itoh et al. of Waseda University Mechanical Engineering department robot research team in Japan have made the condition reflection model using the Hull theory, and it could cause the WE-4RII robot to learn the sentimental expression with the humanity, like shook hand [3]. In 2007, taniguchi

[+] Corresponding author.
E-mail address: st99@heut.edu.cn.

et al. of Japanese Kyoto University proposed condition reflex model depending on plastic peak synchronized signal, and this model can realize to model and the control about the higher order condition reflection, and it could cause the animal to learn the forecast about some important events [4]. This foundation about behavior model is the condition reflex theory, and it has realized some kind of biological characteristic about the robot.

The advanced updating is the rational of establishing the reinforcement learning model, the reinforcement learning utilizes the advanced updating to embody the biological characteristic, but in the behavior choice aspect, it still faces the relational compromise question between exploration and exploitation. In 2002, Dayan et al. in University of London calculated nerve academy of science use the role of the advanced updating in the reinforcement learning to highlight the evolution characteristic of the biological learning, and finally the robot has completed the experiment about the labyrinth successfully [5]. In 2004, Doherty in neurology research institute and Dayan et al. in University of London investigate together, and point out the advanced updating is the foundation of the forecast error signal in the reinforcement learning, through the technology of the functionality nuclear magnetic resonance image formation, it has reflected the correspondence relations between the phenomenon of the reinforcement learning and the partial function of the organism [6]. The above instances manifests that the bionic learning is important for the robot by established the reinforcement learning model with the advanced updating.

But at present, almost nobody unifies the advanced updating and the condition reflection learning mechanism, and carries on the research about self-balance control aspects of the two-wheeled robot. In view of this problem, this article proposed a kind of the reinforcement learning mechanism based on the advanced updating as a self-balance bionic learning algorithm of the robot. It can cause the robot to learn the movement balancing control and display the biological behavior as the human or the animal through interacting with the environment.

2. Structure and Mathematical Model of the Robot System

2.1 Structure of the robot system

This article adopts the self-balance two-wheeled robot, and its kinematic scheme uses a type of coaxial difference drive by the two electromotors, and obtains the wheel rotational speed information of the two wheels by computing through the increamental optical-electrical encoder, and checks the posture information of the system through the attitude sensor in time. The core part of the robot control system is TMS320F2812 DSP processor under the chassis of the robot.

The robot chooses two sensors including an inclination angle sensor and a gyroscope to check the posture of the chassis. Supposing, the angle deviating from vertical direction by the chassis is θ, and its angular speed is $\dot{\theta}$, where, the angle θ is checked by the inclination angle sensor, and the angular speed quantity $\dot{\theta}$ is measured by the gyroscope. The structure drawing of the robot is shown in Figure 1.

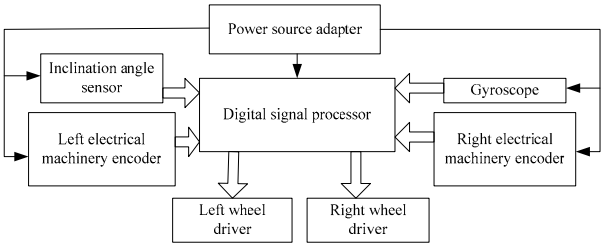

Figure 1. The structure drawing of the robot

2.2 Dynamics model of the robot

The dynamics model of the two-wheeled robot is established according to the Lagrange equation [7]. It indicates that the movement balance control of the robot required four posture information to realize, and those are tilt angle θ, angular speed of the robot body $\dot{\theta}$, and angular speeds of the left wheel θ_l and right wheel θ_r. According to the system dynamic analysis, the state variable is $\boldsymbol{X} = (\dot{\theta}_l, \dot{\theta}_r, \dot{\theta}, \theta)^{\mathrm{T}}$ and the control variable is $\boldsymbol{U} = (u_l, u_r)^{\mathrm{T}}$. So, the state equation of the system is obtained by computing, and the state equation is shown in formula 1.

$$\dot{\boldsymbol{X}} = A\boldsymbol{X} + B\boldsymbol{U} \tag{1}$$
$$\boldsymbol{Y} = C\boldsymbol{X} + D\boldsymbol{U}$$

Linearizing to the above system state equation, namely regarding to the balance point nearby $|\theta \le 10^{\circ}|$, and supposing it is $\sin\theta = \theta$, $\cos\theta = 1$, when the parameter matrixs of this model respectively are:

$$A = \begin{bmatrix} 0.2381 & -0.6734 & 0 & -87.3042 \\ -0.6734 & -0.2381 & 0 & -87.3042 \\ 0.9320 & 0.9320 & 0 & 136.6289 \\ 0 & 0 & 1 & 0 \end{bmatrix};$$

$$B = \begin{bmatrix} -0.1327 & 0.9781 \\ 0.9781 & -0.1327 \\ -2.1384 & -2.1384 \\ 0 & 0 \end{bmatrix};$$

$$C = \begin{bmatrix} -0.139 & 0 & 0 & 0 \\ 0 & -0.139 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix};$$

$$D = 0.$$

3. Design of the Bionic Self-Balance Controller based on the Advantage Learning

Advanced updating embodies the essence character of the reinforcement learning, but the reinforcement learning is an operation skill theory about the creature learning, and it is also one learning form about the creature. It allows the agent to adjust own behavior according to the curiosity mechanism so that it can obtain the optimal behavior.

In order to make the robots complete the balancing control task independently and realize the self-balance control goal like creature at the unknown environment, this article uses a kind of reinforcement learning mechanism based on the advanced updating as a self-balance bionics learning algorithm of the robot. Using this algorithm, the robot may grasp the movement balance control skill like person or the animal through the bionic learning process.

3.1 Structural design of the biological controller

The self-balance bionic learning algorithm of the robot based on the advanced updating is designed according to the reinforcement learning theory. It embodies the evolution characteristic of the biological learning by using the advanced updating, and the curiosity mechanism in this bionic self-learning algorithm adopts Boltzmann machine, so that it can choose a certain behavior in term of the probability. The controller system structure diagram is shown in Figure 2.

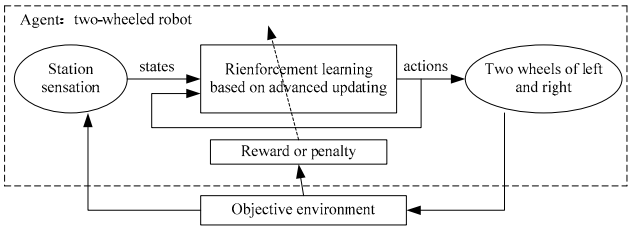

Figure 2. The structure chat of the bionic leaning system of Skinner's operant conditioning mechanism based on advantaged learning.

The bionic self-learning controller consists of the detector device, the controller and the actuator independently. Among them, the detector device can make the robot sense the changes of the external

environment. The controller can improve each kind of control performance according to the measured state variable, and create the satisfied control signal, and the actuator can carry out the corresponding movement according to the actor variable brought by the controller, until the robot learned to control the movement balance of the robot.

3.2 Design of the bionic algorithm

1) Design of the tropism mechanism about probability learning

Advanced updating [8] has provided the explanation for the operant conditioning in two processes. The first part is the evaluation, with a time difference forecast error signal, and it renews the next forecast future reward connected with exterior condition and internal environment (decided by stimulation arrangement). Another part is the behavior, it revises the stimulation-response or the stimulation-response-response association by the strategy, so that the behavior related with the great long-term reward is chosen by a more probability in afterwards experiment. The advanced updating value estimates the behavior; the behavior is chosen completely according to the advance, but the choice relies on the curiosity mechanism.

The curiosity is the habit adapting environment and the survival in the creature evolution process forms. Regarding self-balance robot, its curiosity is the erectness, balanced and stable. In the curiosity mechanism migration probability is obtained based on the Boltzmann machine conformity stimulation probability, and it can cause the station of the system to obey the Boltzmann-Gibbs distribution, and finally tends to thermal equilibrium station [9], namely:

$$P_{TR}(s \to s') = \begin{cases} 1 & \Delta\xi < 0 \\ \exp(\Delta\xi / T) & otherwise \end{cases} \tag{2}$$

When the energy change value is bigger than zero, migration probability P can obey exponential distribution, namely, when the energy change value is bigger than zero, the probability P will decrease, and when it is smaller than zero, the probability P is a zero. This indicated that, in the migration probability of the orientation mechanism reduces along with temperature T, and the probability chosen the poor movement can get smaller, on the other hand, the probability chosen the optimization movement can increase. Just like the physiology description of reinforcement learning, if it the consequence produced is the positive reinforcement signal, then the appearance probability of this behavior in the next time will be increased; otherwise, if it the consequence produced is the negative reinforcement signal, then the appearance probability of this behavior in the next time will be reduced. This just likes the basic principle of the reinforcement learning, and as the curiosity mechanism, the Boltzmann machine realizes the behavior choose characteristic of creature in the process of the reinforcement learning.

2) Design of the bionic learning algorithm based on the advanced updating

The reinforcement learning mechanism based on advanced updating can estimate the behavior according to the advanced value. The big advanced value of the behavior can be first considered. If the advanced value of the behavior is a negative value, it can be abandoned. And if the creature does better, this behavior under this condition will have a higher frequency. The better process of the movement choice is one kind of strategy improvement form, therefore, this article uses Boltzmann machine as the curiosity mechanism about behavior choose probability of the robot.

When some chosen behavior changes, the state value and advanced value also changes. The forecast error signal value of behavior execution in fact is the advanced goal. Therefore, the advanced error is the TD forecast error δ after the state chooses behavior. The advanced error signal can change the weight of the synapse. Among them, the advanced value refers is the movement estimate function J, namely: $J(t) = r(t+1) + \gamma \cdot r(t+2) + \gamma^2 \cdot r(t+3) + \cdots$. And the award signal r is one kind of the function effect appraisal to the action chosen. The balance state of the robot refers that the various states quantities of the robot satisfy that, namely, The inclination angle of the swing link $\theta < 0.0523 rad$, and the angular speed of the swing link $\dot{\theta}$, and the angular speed of the left wheel $\dot{\theta}_l$ and the right wheel $\dot{\theta}_r$ are smaller than $3.489 rad/s$, and the discount factor $\gamma = 0.9$ expresses the important degree of the reward forecast in the near future and specified future. The movement estimate function goal is causes the long-term estimate

function value J to be biggest, and it indicates this behavior can produce the most superior effect. Where, the sampling time will be 0.02s.

When the operations are still under their control, they can achieve the corresponding tasks according their advantage. An essential supposition is that the advantage has a baseline, and when an advantage of the behavior is under the baseline, this behavior can be removed from the competition. This learning mechanism is an online study process, and its environment is unknown, but the current states quantity may be gained.

4. Simulation Experiments and Analysis

In order to confirm the validity of the self-balance robot biological algorithm based on the advanced updating, we made the study simulation experiment about the balancing control of the robot, and gave the corresponding simulation result.

In the simulation experiment, the start value of the robot is stochastic, and it can make the robot start the exploratory study. The result indicates, the robot passes through 173 tests in 172 trails under the ideal situation, and finally it can maintain balance control skill through the unceasing learning. Its state curves are shown in Figure 3.

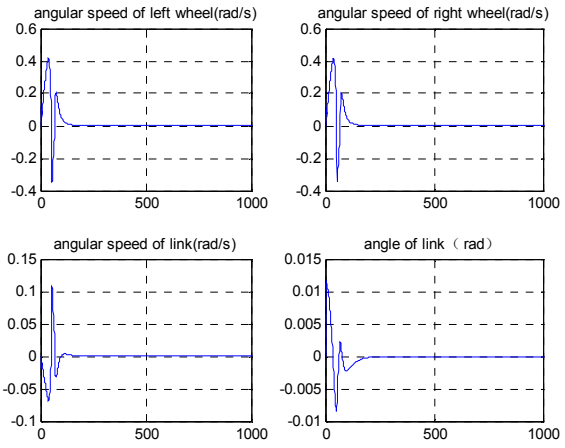

Figure 3. The curve about state variables of the robot.

5. Conclusion

According to the characteristic of the bionics, we proposes the bionic self-balance algorithm based on advanced updating, and the robot has learned the balancing control the skill through interaction with environment. The reinforcement learning is a kind of research about how the creature chooses the behavior, and this behavior is suitable for to express the emotion structure environment. The reinforcement learning permits creature to realize the price expense between response and the reward or the penalty through learning. But the process of the advanced updating manifests the biological characteristic and recombines Boltzmann strategy curiosity mechanism to choose the behavior. And finally, this has realized the integrity reinforcement learning. Namely, the biological movement process can express with a more explicit model, and apply it on the robot, so that the robot can have certain intellectualization characteristic of the creature. It shows that the robot can obtain movement balancing control skill as same as creature and maintain the robot balance through the self-learning and the training under the unknown environment. The bionic algorithm can satisfy the anticipated control goal, and it manifests the bionic self-learning capability of the robot.

6. References

[1] E. Zalama, P. Gaudiano, and J.L. Coronado. Obstacle avoidance by means of an operant conditioning model. From natural to artificial neural computation lecture notes in computer science, 930, 1995, pp471-477.

[2] P. Gaudiano, C. Chang. Adaptive obstacle avoidance with a neural network for operant conditioning: Experiments with real robots. 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA 97) - Towards New Computational Principles for Robotics and Automation, JUL 10-11, 1997, pp13-18.

[3] K. Itoh, and et al. Behavior model of humanoid robots based on operant conditioning. 2005 5th IEEE-RAS International Conference on Humanoid Robots, v 2005, 2005, pp.220-225.

[4] T. Tadahiro, S. Tetsuo, Incremental acquisition of behaviors and signs based on a reinforcement learning schemata model and a spike timing-dependent plasticity network. Advanced Robotics, V.21, N.10, 2007, pp1177–1199.

[5] P. Dayan, BW. Balleine, Reward, motivation, and reinforcement learning——Review. NEURON. v:36, N:2, 2002, pp. 285-298.

[6] O'D. John, and et al. Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. Science, 2004, pp. 451-454.

[7] Zhang Xiaohua, System model and simulation. Tsinghua University Pres, 2006, pp. 224-232.

[8] O'D. John, D. Peter, S. Johannes, et al, Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. Science, 2004, pp. 304-452.

[9] Y. Dahmani and A. Benyettou, Seek of an Optimal Way by Q-Learning. Journal of Computer Science 1, 2005, pp. 28-30.

[10] R. Hongge, R. Xiaogang, A Bionic Learning Algorithm Based on Skinner's Operant Conditioning and Control of Robot 32, 2010, pp. 132-137.