ICIC Express Letters
Part B: Applications
Volume **7**, Number **11**, November **2016**
 ISSN 2185-2766
pp. **2291–2300**

RESEARCH ON TWO-WHEELED SELF-BALANCE ROBOT BASED ON IM-Q-ELM ALGORITHM

Hongge Ren, Rui Yin, Tao Shi and Fujin Li

College of Electrical Engineering
North China University of Science and Technology
No. 46, Xinhua Road, Tangshan 063009, P. R. China
384042235@qq.com

Received May 2016; accepted August 2016

Abstract. *In view of the problem of the poor initiative in the balance of motion control of two-wheeled self-balance robot, an autonomous learning algorithm (IM-Q-ELM algorithm) based on intrinsic motivation in extreme learning machine, which is inspired by the theory of intrinsic motivation in psychology, has been proposed. Firstly, the algorithm replaced internal reward signal by intrinsic motivation mechanism, simulated cognitive mechanism of human's brain, and improved the independent performance of learning. Secondly, we used extreme learning machine network to save knowledge and experience and let robots learn balance skills gradually. At last, the independence and rapidity of the algorithm were verified by simulation experiments. Simulation experimental results show that this algorithm has better self-balance control skill, good robust performance and high practical application value.*
Keywords: Two-wheeled self-balance robot, Intrinsic motivation, Reward mechanism, Extreme learning machine, Robustness

1. **Introduction.** Nowadays, with the development of intelligent technology, robot technology plays an extremely important role in people's production and life. It not only can replace human beings to complete some relatively heavy tasks, but also can improve the work efficiency to a certain extent. It also saves a lot of cost and manpower. In recent years, more and more experts and scholars put forward some control methods for the control of two-wheeled self-balancing robot. In 2006, Zhang and Wu combined Q-learning algorithm with BP neural network to achieve the status information without discretization of the inverted pendulum model learning control, and improved the learning speed [1]. In the same year, Pfeifer and Bongard stressed that robot learned in unknown environment through interacting of robot and environment [2]. In 2010, Ren and Ruan adopted self-regression neural network learning algorithm based on Skinner operating condition reflection theory to be the learning mechanism of the robot. It not only completed the control of two-wheeled self-balancing robot, but also verified the robustness of this algorithm at the same time [3]. In 2013, Cederborg and Oudeyer combined the thought of intrinsic motivation (IM) with exploration problem of biological self consciousness, and proposed a status transition error learning machine of the system. It achieved learning actively of the robot based on intrinsic motivation model in an unknown environment [4]. In 2014, Hu and Qu proposed an autonomous learning method, which was driven by intrinsic motivation in unknown environment. It was inspired by the intrinsic motivation of psychology. It improved the convergence of the algorithm and decreased the error of system. The degree of intelligence also improved significantly [5]. In 2014, Wang et al. proposed a short term power load model for online sequential extreme learning machine (OSELM algorithm). Its accuracy was better than that of generalized neural network and SVM, and it had a good performance in parallel [6]. In 2015, Ren et al. proposed a

reinforcement learning algorithm based on intrinsic motivation. It used the intrinsic motivation signal as the internal reward, then simulated human psychological mechanism, and applied to the whole learning process with the external signal. It not only ensured the rapidity of the system, but also improved the self-learning ability greatly [7]. However, there are a lot of limitations of slow learning speed, poor stability and robustness in existing simple methods such as BP network, and PID model. And the proposed method solves above-mentioned problems.

In view of the research on controlled problem of two-wheeled self-balance robot, an autonomous learning algorithm (IM-Q-ELM algorithm) based on intrinsic motivation in extreme learning machine has been proposed in this paper. This algorithm takes Q-learning as a frame and puts the intrinsic motivation signal as an intrinsic incentive to drive the robot learning. At the same time, the extreme learning machine network is used as the storage space of knowledge accumulation. It makes robot like a person, self-learning, self-organization, and gradual formation. It also improves the balance control skill through the imitation of human brain learning model.

In Section 1, we introduce some scholars' research results in recent years, illustrate the limitations of these methods and describe the superiority of proposed algorithm. In Section 2, we build two-wheeled robot model according to Lyapunov model and explain the model vectors. In Section 3, we design the IM-Q-ELM algorithm and explain the eight-tuple computing model. And in Section 4, we verify the effectiveness and rapidity of the proposed algorithm. Finally, Section 5 concludes this paper and our ideas for future work.

2. Structure and Dynamic Model of Two-wheeled Self-balance Robot System.

Two-wheeled self-balance robot [8] is designed by the typical inverted pendulum model. It has only two wheels, and arranges different speed in two wheels. Each wheel is driven by a DC motor through a reducer directly. They rotate around the axis of the motor through the center. The robot's center of gravity is above the two axles. The robot maintains the dynamic balance by the movement. The system structure diagram is shown in Figure 1.

In order to control and analyze two-wheeled self-balance robot accurately, firstly, the mathematical model of two-wheeled self-balance robot is established [9]. Secondly, we have neglected the influence caused by uneven pavement in practical environment, wind resistance and so on. Figure 2 is the simplified structure of two-wheeled robot system. The meanings of parameters are shown in Table 1.

3. An Autonomous Learning Algorithm Based on Intrinsic Motivation.

3.1. Designing the algorithm.

The activities of human and animal organisms can produce signals in different regions of the brain through the external and the internal environments. Figure 3 is the structure of an autonomous learning algorithm based on

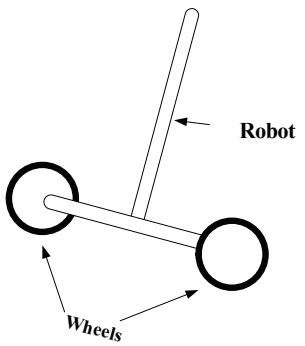

FIGURE 1. Structure of two-wheeled robot system

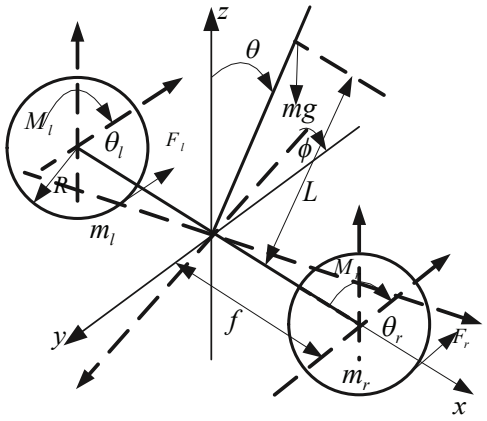

FIGURE 2. Simplified structure of two-wheeled robot

TABLE 1. Parameters of two-wheeled robot

Symbol	Name
θ	The angle between z axis and body
ϕ	Turning angle of robot body
L	Distance between the center of mass and the axle
R	Radius of wheels
F_l	Friction of left wheel
F_r	Friction of right wheel
$2f$	Length between two wheels
θ_l, θ_r	Rotation angle of left and right wheels
m_l, m_r	Weight of a wheel
M_l, M_r	Motor torques
m	Weight of robot
V_l, V_r	Velocities of two wheels
V_x, V_y, V_z	Velocities on x, y, z axes
J_φ	Movement of inertia rounded by z axis
J_θ	Movement of inertia rounded by x axis
J_l, J_r	Movement of inertia of two wheels

intrinsic motivation. Firstly, the robot obtains the status input information value from external environment, and then calculates the reward value according to the intrinsic motivation orientation function. We can learn and train about its status information through extreme learning machine neural network and put current environment status information in the network node to be the experience value of next action decision. At the same time, the system outputs the corresponding action decision, and explores the unknown environment ulteriorly.

An autonomous algorithm based on intrinsic motivation can be represented as an eight-tuple computing model:

$$IM\text{-}Q\text{-}ELM = \{S, A, f, H, R(s_t, a_t), Q(s_t, a_t), V(t), p(s_t, a_t)\}$$

The meaning of each element is as follows.

(1) S: It is the internal status set of IM-Q-ELM algorithm. $S = \{s_i | i = 1, 2, 3, \ldots, n\}$, where s_i represents the ith status, and n represents the number of all generating status information.

(2) A: It is the action set of IM-Q-ELM algorithm. $A = \{a_i | i = 1, 2, 3, \ldots, m\}$, where a_i represents the ith action, and m represents the number of all actions.

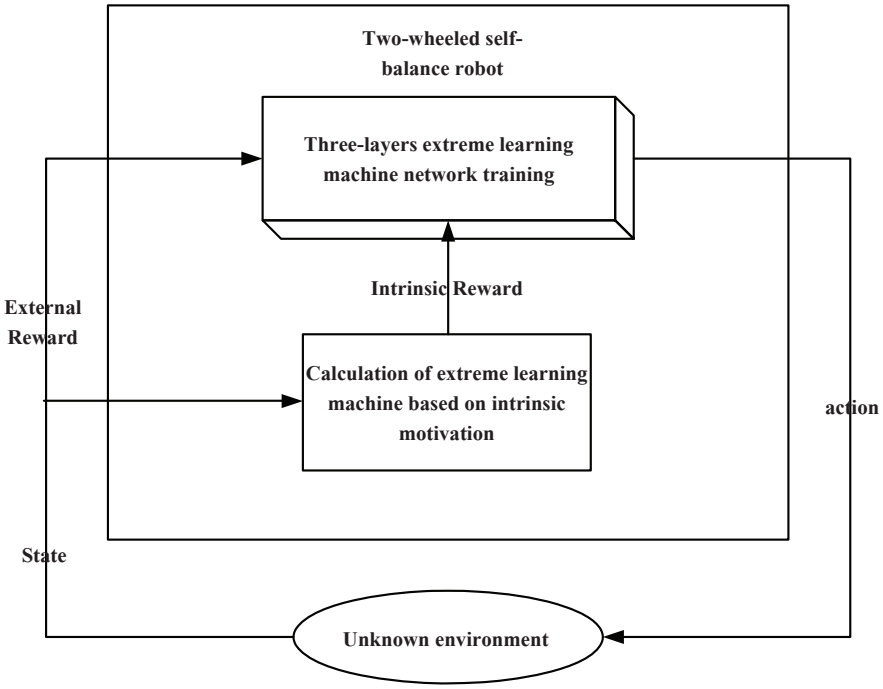

FIGURE 3. Structural framework of autonomous algorithm based on intrinsic motivation

(3) f: It is the status transition function of IM-Q-ELM algorithm. Generally speaking, it is determined by the system model and the environment.

(4) H: It is the intrinsic motivation orientation function of IM-Q-ELM algorithm. Intrinsic motivation in psychology is described in terms of strange degrees, curiosity, boredom degrees and so on. It is the driving force for the exploration and study of drivers or other living creatures. This driving force can be attributed to the orientation mechanism function introduced by the intrinsic motivation in the learning process. We will simulate the working mechanism of the human brain. So that it has the ability of independent learning and improves learning efficiency. The intrinsic motivation mechanism of the system is

$$H(t) = \frac{1 - e^{-\lambda V(t)}}{1 + e^{-\lambda V(t)}} \tag{1}$$

Among them, λ represents the parameter of orientation function. The less value of $H(t)$ is, the less intrinsic motivation orientation of the system is. It indicates that the corresponding action reward is less. On the contrary, the greater the corresponding action reward is, the greater the intrinsic motivation orientation of the system is.

(5) $R(s_t, a_t)$: It is the reward signal of IM-Q-ELM algorithm. At the time of t, status of s_t, the reward value of the system is the value at the status of s_{t+1} after executing the action of a_t.

(6) $Q(s_t, a_t)$: It is the iterative formula of reinforcement learning of IM-Q-ELM algorithm.

(7) $V(t)$: It is the evaluation function of IM-Q-ELM algorithm.

(8) $p(s_t, a_t)$: It is the action selection probability of IM-Q-ELM algorithm.

In this algorithm, the intrinsic reward function is replaced by the intrinsic motivation. We assume that the agent is running in a strange environment. Its output is x_t, and the desiring output is defined as $\hat{x}_t$. And then the difference value ($r_{in} = x_t - \hat{x}_t$) between them is defined as the system's internal reward function. When the system selects the action a at time of t, the status will be transferred from s_t to s_{t+1}. If $r_{in}(t+1) - r_{in}(t) < 0$, that is, error of the system is smaller than error of the previous time, it shows the expected effect of target status that action selects at the time of $t+1$ it is better than the expected

effect of target status that the action selects at the time of t. At the same time, it also shows that the system has greater orientation in time of $t+1$. On the contrary, if $r_{in}(t+1) - r_{in}(t) > 0$, the system has smaller orientation in time of $t+1$.

We believe that the evaluation function ($V(t)$) of the internal action is approaching to zero gradually under the driving mechanism of motivation. So that the two-wheeled robot can maintain the most appropriate balance, and we define the evaluation function:

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

where $0 \leq \gamma \leq 1$ is the discount factor.

The evaluation function at time of $t-1$ is:

$$\begin{aligned} V(t-1) &= r(t) + \gamma r(t+1) + \gamma^2 (t+2) + \cdots \\ &= r(t) + \gamma \left[r(t+1) + \gamma r(t+2) + \cdots \right] \\ &= r(t) + \gamma V(t) \end{aligned} \tag{3}$$

This proves that the evaluation function in time of $t-1$ can be expressed by the evaluation function in time of t. We can use $V(t)$ as an observer to observe $V(t-1)$. And then we can establish a TD error difference formula as follows:

$$\varphi = r(t) + \gamma V(t) - V(t-1) \tag{4}$$

3.2. **Autonomous learning model.** Based on the Q-learning framework, it combined with the thought of intrinsic motivation to drive agent, and put the intrinsic motivation orientation function as an incentive mechanism of the algorithm. Then it trains with extreme learning machine network to increase autonomous learning ability of robot and to improve the learning speed greatly. The control structure diagram of the algorithm is shown as Figure 4.

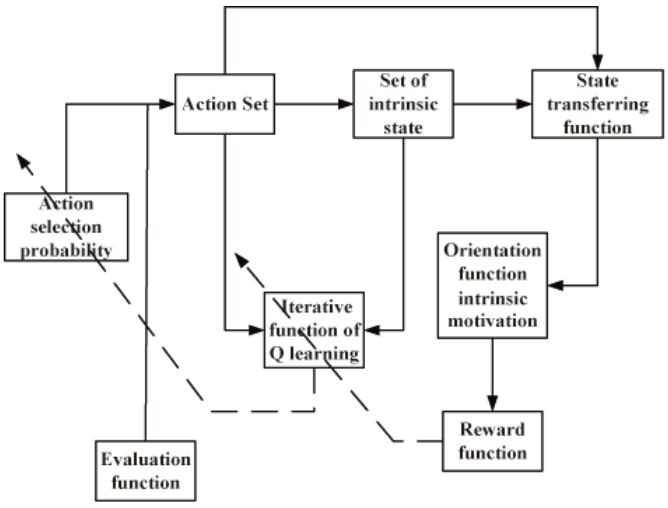

FIGURE 4. Extreme learning machine autonomous learning algorithm structure based on intrinsic motivation

In the classical Q-learning algorithm, it calculates iteratively according to Markov decision process behavior value function through time error TD algorithm. The iterative formula is:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \kappa \left[R(s_t, a_t) + \gamma_{\max} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t) \right] \tag{5}$$

where $\kappa > 0$ is learning factor.

The update of the reward signal is driven by intrinsic motivation.

$$R = \xi r^{in} + \eta r^{ex} = \xi H + \eta r^{ex} = \xi \frac{1 - e^{-\lambda V(t)}}{1 + e^{-\lambda V(t)}} + \eta r^{ex} \tag{6}$$

In the formula, r^{in} represents intrinsic motivation function, r^{ex} represents external motivation function, and ξ and η represent the weight of r^{in} and r^{ex} respectively. So the iterative formula is:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \kappa \left[\left(\xi \frac{1 - e^{-\lambda V(t)}}{1 + e^{-\lambda V(t)}} + \eta r^{ex} \right) + \gamma_{\max} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t) \right] \quad (7)$$

3.3. **Algorithm process.**

Step 1: Initialize the current status s_0 of the system. Select the discount factor $\gamma = 0.86$ and learning factor $\kappa = 0.92$. And select the appropriate intrinsic motivation function and the weight of the external motivation function ξ and η;

Step 2: Calculate all the Q value of possible actions;

Step 3: Select the appropriate action according to the Q value;

Step 4: Execute the current action, and make decisions for the next learning phase;

Step 5: Calculate intrinsic motivation function H. At the same time, calculate the optimal action decision according to the reinforcement learning;

$$\hat{a}_t = \arg\max \left[Q_{a_t \in A}(s_{t+1}, a_{t+1}) \right] \quad (8)$$

Step 6: Update Q value according to Formula (7);

Step 7: Update current time $t = t + 1$, and current status $s_t = s_{t+1}$;

Step 8: Repeat **Step 2**∼**Step 7** until the training completed.

4. Experiment Results and Simulation.

4.1. **Experimental designing.** We put two-wheeled robot completing self balancing in the unknown environment as the control object. In the self learning mechanism of extreme learning machine based on intrinsic motivation, $Net_3(4, 8, 1)$ was selected by action extreme learning machine. The four status inputs respectively presented the robot's own angle, its own angle rate, cart position and cart rate. Output was the control quantity of two-wheeled robot. Evaluation network selected $Net_3(5, 8, 2)$. The input variables were the four states of the robot and the control quantity of the action extreme learning machine which outputted from the network output layer. The outputs were evaluation function $V(t)$ and the driving force $F(t)$ variation of the robot body.

$$V(t) = r(t + 1) + \gamma r(t + 2) + \gamma^2 r(t + 3) + \cdots \quad (9)$$

Among them, r was the reward value. When the robot's own inclination $\theta < 0.025$rad, the system would get a reward value $r = 0$; otherwise $r = -1$. In Section 3.3, we selected the appropriate discount factor and learning factor. Sampling time selects $T = 0.01$s. The probing time of each experiment was more than 200 times, or the number of steps in the training process was more than 15000 steps. So we thought the experiment failed. At this point, we should terminate the test, and retest. If robot could maintain the 15000 steps and not fall down once in an experiment, it indicated that robot completed the balance controlling in an unknown environment. After each test failure, the initial state, weight value and threshold value were reset in a certain range of random values, and trained again. The results of the 60 experiments were summarized that robot could achieve self-balancing control after 65 failures. It reflected the strong autonomous learning and adaptive ability. The simulation results were shown in Figure 5.

4.2. **Analysis of experimental results.** In order to verify the effectiveness and convergence of the proposed algorithm, simulation experiments were carried out on the performance of the two-wheeled robot system, and analyzed the experimental results.

Figure 5 shows four states' variables with time curve of two-wheeled self-balance robot which is trained through the autonomous learning algorithm of extreme learning machine based on intrinsic motivation (IM-Q-ELM algorithm) mentioned in this paper. From the figure, we can see that robot can complete the self balance control after 3s (that

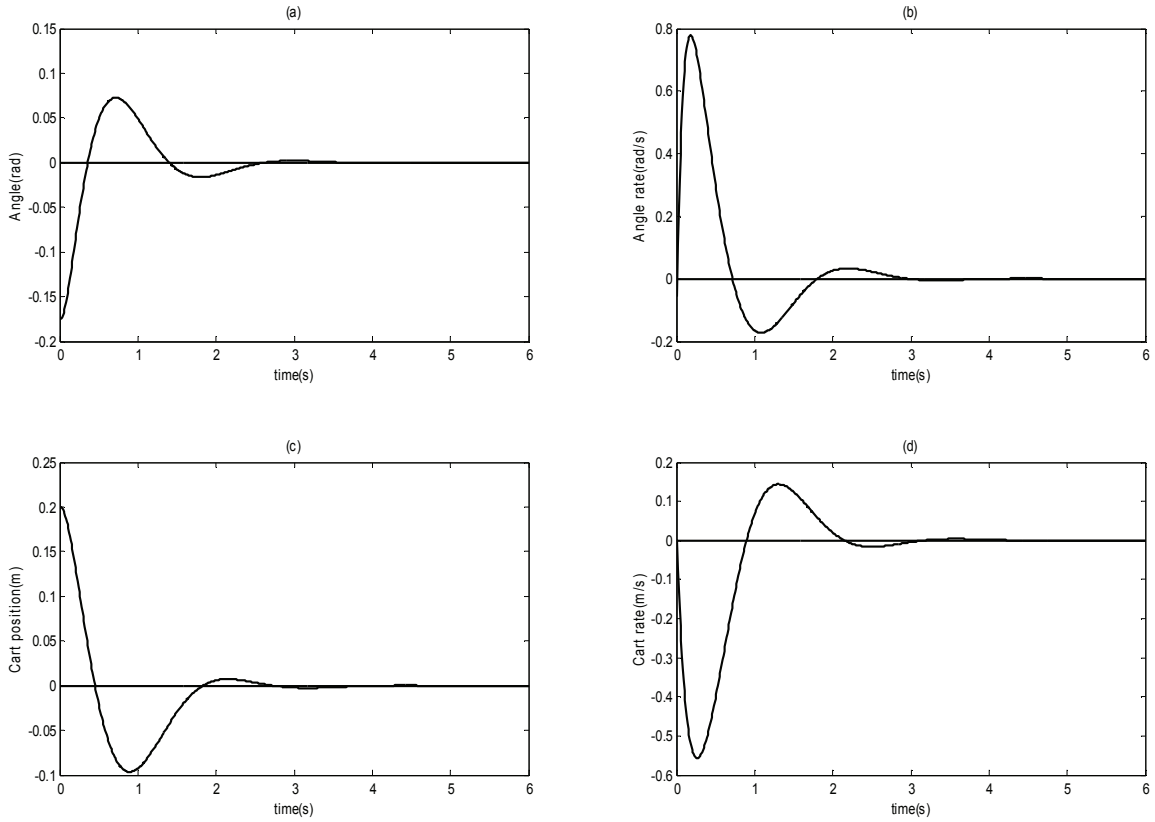

FIGURE 5. Curve of states' variables

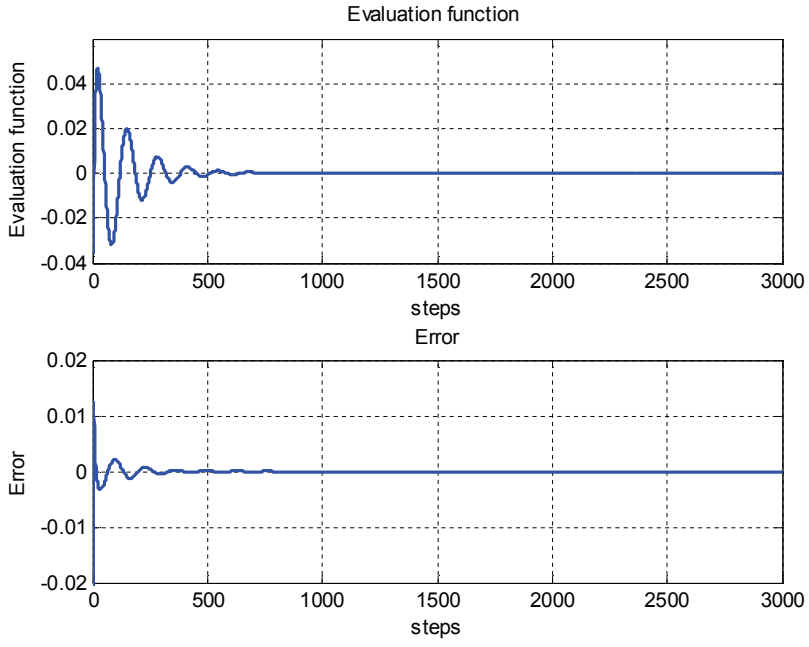

FIGURE 6. Curve of evaluation function and error

is 300 steps). It shows fast autonomous learning and adaptive ability of the algorithm. Figure 6 shows the evaluation function curve and error curve of system states of the robot by 3000 steps in the training process. Figure 7 is the curve of robot force change. Figure 8 shows the evaluation function simulation comparison between the autonomous learning algorithm of extreme learning machine based on intrinsic motivation (IM-Q-ELM algorithm) and traditional reinforcement learning algorithm (RL). Figure 9 shows the error

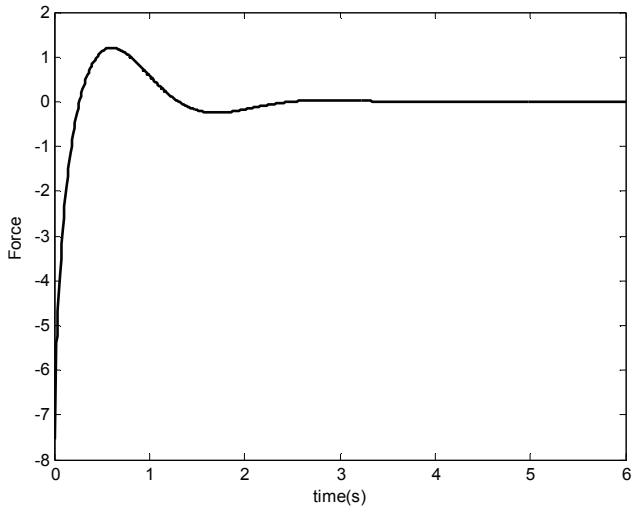

FIGURE 7. Force curve of robot

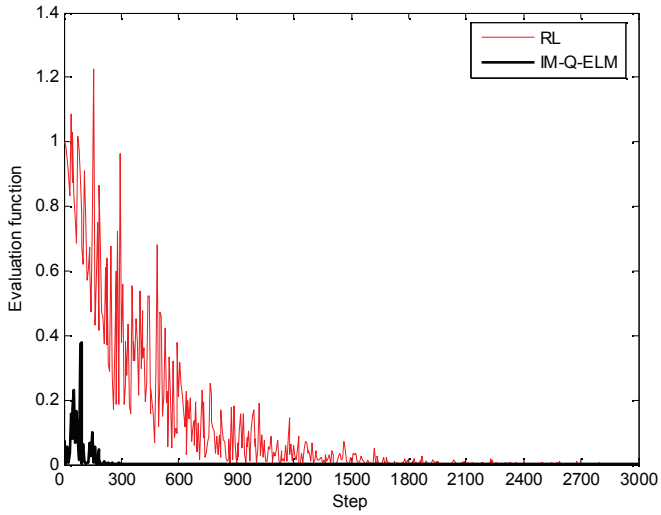

FIGURE 8. Comparison of evaluation function simulation

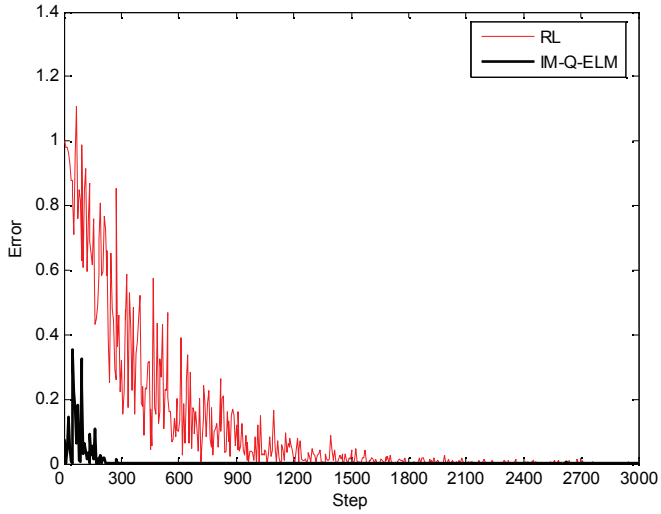

FIGURE 9. Comparison of error simulation

simulation comparison between above two algorithms. It can be seen that the robustness of IM-Q-ELM algorithm proposed in this paper is far more than the latter. As a result, we can learn from the above simulation experiments that reinforcement learning with intrinsic motivation can get better performance and faster learning and training speed after the training of extreme learning machine network. At the same time, it also shows the strong autonomous learning ability and control ability of two-wheeled self-balance robot.

Table 2 compares the performance of the system with the data obtained in the literature. From the table, we can also see that the autonomous learning algorithm of extreme learning machine based on intrinsic motivation is the fastest, and the training accuracy is the highest.

Table 2. Performance comparison of different learning algorithms

	Attempt times	Balance time	Odds of success
Algorithm in this paper	15000	About 3s	100%
Algorithm in Literature [8]	10000	About 4.5s	89%
Algorithm in Literature [10]	1000	About 9s	97%

5. **Conclusion.** An autonomous learning algorithm (IM-Q-ELM algorithm) for extreme learning machine based on intrinsic motivation was proposed in this paper. It applied to the balance control of two-wheeled self-balance robot. We replace traditional reinforcement learning incentive mechanism by intrinsic motivation mechanism, and determine the value of the evaluation. We can also replace the traditional self organizing neural network by extreme learning machine network. It can greatly improve the speed of agent to adapt to the unknown environment, and make robot learn self balance with relatively short period of time. The simulation experimental results show that the algorithm has strong robustness and fast convergence performance. It also reflects the strong autonomous learning ability of the algorithm. However, the rapidity and robustness of extreme learning machine are quite advantageous. It can also be applied to the fields of image processing, temperature detection and so on. It has larger research space.

Acknowledgement. This work was supported by the National Natural Science Foundation (NNSF) under Grant of China 61203343.

REFERENCES

[1] T. Zhang and H. S. Wu, Balance of an inverted pendulum using neural network and Q-learning, *Computer Simulation*, vol.23, no.4, pp.298-300, 2006.

[2] R. Pfeifer and J. C. Bongard, *How the Body Shapes the Way We Think: A New View of Intelligence (Bradford Books)*, The MIT Press, Cambridge, vol.7, pp.110-111, 2006.

[3] H. G. Ren and X. G. Ruan, Self-banlance control of two-wheeled robot based on Skinner's operant conditioned reflex, *Control Theory & Applications*, vol.27, no.10, pp.1424-1428, 2010.

[4] T. Cederborg and P. Y. Oudeyer, From language to motor gavagai: Unified imitation learning of multiple linguistic and nonlinguistic sensorimotor skills, *IEEE Trans. Autonomous Mental Development*, vol.5, no.3, pp.222-239, 2013.

[5] Q. X. Hu and X. Y. Qu, Internal motivation driven robot online and autonomous learning of unknown environment, *Computer Engineering and Applications*, vol.50, no.4, pp.110-113, 2014.

[6] B. Y. Wang, S. Zhao and S. M. Zhang, A distributed load forecasting algorithm based on cloud computing and extreme learning machine, *Power System Technology*, vol.38, no.2, pp.526-531, 2014.

[7] H. G. Ren, Y. F. Xiang, F. J. Li and W. M. Liu, Research on reinforcement learning algorithm based on intrinsic motivation for two-wheeled robot, *Computer Measurement & Control*, vol.23, no.9, pp.3185-3187,3191, 2015.

[8] X. G. Ruan, J. X. Cai and J. Chen, Learning to control two-wheeled self-balancing robot using reinforcement learning rules, *Computer Measurement & Control*, vol.17, no.2, pp.321-323, 2009.

[9] X. H. Zhang, *System Modelling and Simulation*, Tsinghua University Publisher, Beijing, pp.225-232, 2006.

[10] G. F. Jiang and C. P. Wu, Learning to control an inverted pendulum using Q-learning and neural networks, *Acta Automatica Sinica*, vol.24, no.5, pp.662-666, 1998.