Autonomous Navigation Strategy in Mobile Robot

Jianxian Cai
Institute of Disaster Prevention / Department of Disaster Prevention Instrument, Sanhe Hebei, China
Email: cjxlaq@163.com

Lixin Li
Institute of Disaster Prevention / Department of Disaster Prevention Instrument, Sanhe Hebei, China
Email: lilixin4job@163.com

***Abstract*—To solve the navigation problem of mobile robot in unknown environment, a navigation scheme based on bionic strategy was proposed, which simulates operant conditioning mechanism. In this scheme, the tendency cell was designed by use of information entropy, which represents the tendency degree for state. The improved Q learning algorithm used as learning core to direct the learning direction. The Boltzmann machine was used to process annealing calculation, which can randomly selected navigation action. The selected strategy of action will tend to optimal with the learning process. Simulation analyses were carried out in mobile robot; results showed that the proposed method had quick learning velocity and accurate navigation ability, and robot could successfully evade obstacles and arrived at goal point with optimal path.**

***Index Terms*—mobile robot, navigation, bionic strategy, information entropy, *Q* learning, Boltzmann machine**

I. INTRODUCTION

Robot autonomous navigation [1-5] is one of the key technology and difficulty problems in mobile robot research field. Robot autonomous navigation problem includes environment sense, dynamical decision making, and actions control and so forth. With the expansion of application in many important fields and the development of artificial intelligent technology, mobile robot navigation increasingly develops to the direction of intellectualized and autonomy-oriented. Imitate behavior way and learning ability of human is the developing direction of intelligent robot research.

Reinforcement learning [6-8] and operant conditioning learning [9, 10] preferably accord with psychology habit of solving problem in human, which have received widespread attention in robot study. More importantly, Reinforcement learning and operant conditioning learning have a character that independent of the environment model and learning on-line, so they are considered as a promising machine learning strategy, especially in the unknown environment navigation. Reinforcement learning had been shown to be useful for solving autonomous navigation problems [11]. Of course, this is an effective method, but the information processing velocity is slow and the capacity is limited. Mentionable, operant conditioning learning can explain the details of experiment phenomenon, in which the agent can learn more complicated action. So operant conditioning learning attracts more and more researchers. Operant conditioning learning model had achieved some progress in recent years and had applied to robot motion control [12-16] successfully, however seldom found in robot navigation control.

Based on above ideas, for solving the navigation problem of mobile robot in unknown environment, a navigation scheme based on the bionic strategy according is proposed. This scheme simulates operant conditioning mechanism, in which, the tendency cell is designed by use of information entropy which represents the tendency degree for state. The improved Q learning algorithm used as learning core to direct the learning direction. The Boltzmann machine is used to process annealing calculation, which can randomly select navigation action. The selected strategy of action will tend to optimal with the learning process. Simulation analyses are carried out in mobile robot; results show that robot can accomplish

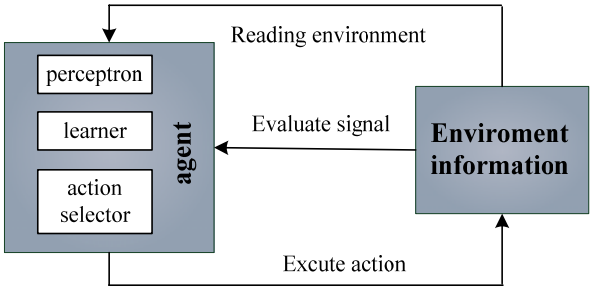


Figure 1. Model of operant conditioning learning.

Manuscript received June 14, 2012; revised July 3, 2012; accepted in 2012.
Project number: Special Fund of Fundamental Scientific Research Business Expense for Higher School of Central Government (Projects for creation teams) (ZY20110104) , Teachers' Scientific Research Fund of China Earthquake Administration (20110122).
Corresponding author: Jianxian Cai.

© 2013 ACADEMY PUBLISHER
doi:10.4304/jcp.8.8.2118-2125

the autonomous navigation task by interacting with environment.

II. Implement Operant Conditioning Mechanism

Operant conditioning learning mechanism emphasize on the interaction between agent and environment. The leaning process is shown in Fig.1.

Operant conditioning learning model consist of three modules: perception which responsible for obtaining environment state; learner which responsible for learning; action selector which responsible for selecting suitable action from action space. Robot continuously perceives environment and implements selected action, accordingly the environment will be changed. After that, the evaluate signal of action can be attained according to the change of environment state after implementing the selected action. The learning process will finish after the above process proceeds repeatedly. It can be seen from operant conditioning learning process that there are similar aspects in the process of leaning and adapting environment between higher animal and robot. So it is feasible to study autonomous navigation of robot by simulating operant conditioning mechanism.

III. Implement Autonomous Navigation Algorithm

As shown in Fig.2, this is the mobile robot autonomous navigation diagram by simulating operant conditioning mechanism. State editor module solve the problem of dividing environment states; navigation learning strategy module solve the problem of mapping from state space to action space; environment interaction module solve the problem of produce of evaluate signal and state transfer. Mobile robot keep interact with environment in the process of navigation learning and will achieve anticipative object via learning and training.

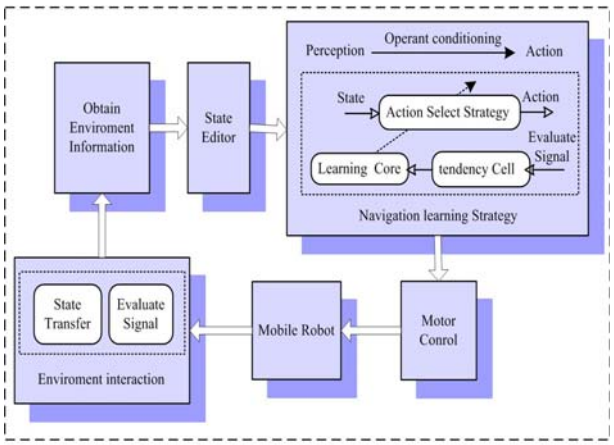


Figure 2. Navigation strategy system of simulating operant conditioning.

A. Divide Environment State

Robot continuously learns in the process of navigation, and environment state will transfer after learning. Learning produce new interaction, and new interaction will stimulate further learning. Robot action will become better and better and gradually adapt the environment by repeated iteration. System state should be decomposed and form state space for promoting environment interaction.

State and action of mobile robot in local unknown environment are considered, and the navigation aim is that robot can arrive goal point from start point in collision-free. The relationship schematic diagram between robot and environment is shown in Fig.3. Robot equips with sixteen sonar sensors, and the distance among sonar sensor is 20° or 40°. The coverage area of sonar sensors contain 0 ～ 360° and so robot can freedom rotate in the area of 0～360°.

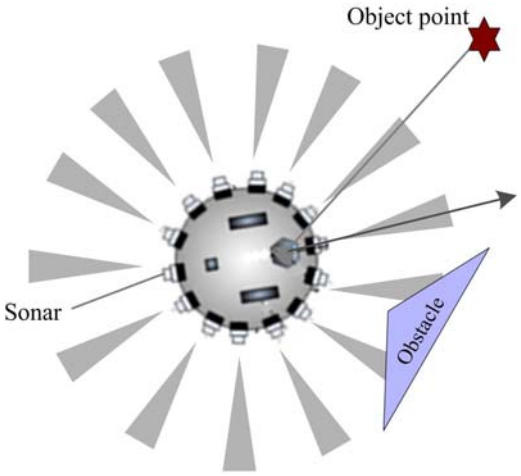


Figure 3. Relationship schematic diagram between robot and environment.

Suppose robot can continuously motion with equal space and each step size is a grid length for conveniently study navigation arithmetic. At the same time, robot can freedom turn round with collision-free in narrow environment area. So robot is simplified as a particle and not considers the rotation radius of robot in navigation arithmetic. The relation between robot and obstacle object is shown in Fig.4.

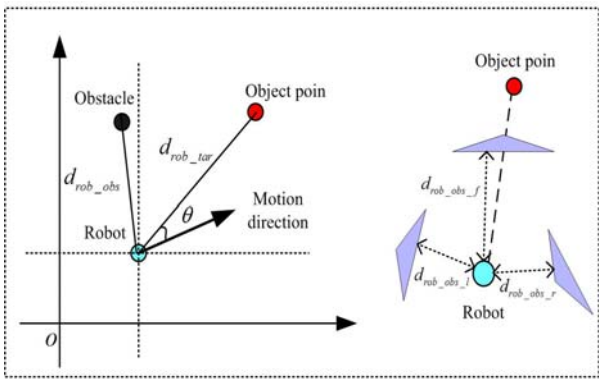


Figure 4. Relation between robot and obstacle.

© 2013 ACADEMY PUBLISHER

The detection range of sensor is divided in to three area (left, frontage and right) for reduce state space, in addition, considering the relation between robot and object, the definition of state space is $s=\left\{d_{rob\_obs\_l},d_{rob\_obs\_f},d_{rob\_obs\_r},d_{rob\_tar},\theta\right\}$, where $d_{rob\_obs\_l}$ is the distance between robot and left obstacles, $d_{rob\_obs\_f}$ is the distance between robot and front obstacles, $d_{rob\_obs\_r}$ is the distance between robot and right obstacles, d_{rob_tar} is the distance between robot and target, θ is the angle between robot motion direction and target.

The sense range of sonar sensor in robot is 100mm~5000mm, so $d_{\min}=0.1\text{m}$ is defined as the least dangerous distance and $d_{\max}=5\text{m}$ is defined as safe distance. The angel range between robot motion direction and target is $\theta\in\left[-180^{\circ},180^{\circ}\right]$. The state variable is divided in to five levels (see TABLE I)

B. Define Action Space

The definition of action space should follow two principles. The first is that there are enough actions to accomplish robot navigation task, and the second is that the actions should be retrench for avoiding learning burden. So action space defined as $A=\left\{a_1,a_2,a_3,a_4,a_5\right\}$, where a_1 means left turn 30° and move 0.1m, a_2 means left turn 10° and move 0.1m, a_3 means turn 0° and move 0.1m, a_4 means right turn 10° and move 0.1m, a_5 means right turn 30° and move 0.1m.

C. Designation of Tendency Cell

Complexity of bionic autonomous learning strategy will index increased with the increase of states number. The related literature pointed out that the related states to best effect are lower than quarter. So information entropy is introduced, which can weigh the tendency of state by computing entropy value of environment state. Entropy can measure uncertainty degree of an event in information theory. The information is bigger the system structure is more regular and the function is more perfect, correspondingly the entropy value is more small. Suppose the discrete environment state set is $S=\{s_i\,|\,i=1,2,\cdots,n\}$, then the information entropy of state $s_i\in S$ is defined as

$$H_i\left(A(t)\,|\,s_i\right)=-\sum_{k=1}^{r}p(a_k\,|\,s_i)\log_2 p(a_k\,|\,s_i) \tag{1}$$

Where $a_k\in A=\{a_k\,|\,k=1,2,\cdots,r\}$ is the kth selected action; A is the selected action set, r is the number of selected action; $p_{ik}=p(a_k\,|\,s_i)\in P_i$ represents the probabilistic value of operant action a_k with state s_i and also called as excited probability value of "state s_i -action a_k " pare, which satisfies $0<p_{ik}<1$, $\sum_{k=1}^{r}p_{ik}=1$; $P_i=\{p_{i1},p_{i2},\ldots,p_{ir}\}\in P$, P_i means the probability vector of the ith state and P represent the overall probability vector.

Robot is blindfold in early navigation and so the probability value of each early action is similar, correspondingly information entropy value of each state is the biggest. The action probability of undergo state will change with the learning process and the corresponding information entropy of state will decrease. The number of undergoing is more and the reducing of information is lower, which show robot is more inclined to the state. The maximal information entropy will be deleted.

TABLE I.
TABLE OF STATE SPACE DISCRETE DIVISION

	Very small	Small	Middle	Big	Very big
d_{rob_obs}	(0.1,0.5)	(0.5,1.5)	(1.5,2.5)	(2.5,4)	(4,5)
d_{rob_tar}	(0,0.5)	(0.5,2)	(2,4)	(4,6)	(6,∞)
$\lvert\theta\rvert$	(0,30)	(30,70)	(70,100)	(100,140)	(140,180)

D. Designation of Learning Core

Robot has two goal in the process of unknown environment navigation, namely obstacle avoidance and approaching goal, so the design of evaluate signal should consideration to several aspect:

① $d_{rob\_obs}>d_{\max}$

Robot turns to the goal direction

$$\begin{cases}\Delta\theta(t)=\theta(t+1)-\theta(t)\\ \Delta\theta(t)<0\end{cases} \tag{2}$$

Robot is expected to approach the goal

$$\begin{cases}\Delta d_{rob\_tar}(t)=d_{rob\_tar}(t+1)-d_{rob\_tar}(t)\\ \Delta d_{rob\_tar}(t)<0\end{cases} \tag{3}$$

Then the evaluate signal can be written as

$$\begin{aligned}V(d_{rob\_tar},\theta)=&-\beta_1\text{sign}\left(\Delta d_{rob\_tar}(t)\right)\Delta^2 d_{rob\_tar}(t)\\ &-\beta_2\text{sign}\left(\Delta\theta(t)\right)\Delta^2\theta(t)\end{aligned} \tag{4}$$

Where β_1 and β_2 are weight coefficient, $0<\beta_1,\beta_2<1$.

② $d_{\min}<d_{rob\_obs}<d_{\max}$

Robot is expected to avoid obstacle

$$\begin{cases}\Delta d_{rob\_obs}(t)=d_{rob\_obs}(t+1)-d_{rob\_obs}(t)\\ d_{rob\_obs}(t)>0\end{cases} \tag{5}$$

Then the evaluate signal can be written as

$$\begin{aligned}V(d_{rob\_obs},d_{rob\_tar})=&\beta_3\text{sign}\left(\Delta d_{rob\_obs}(t)\right)\Delta^2 d_{rob\_obs}(t)\\ &-\beta_4\text{sign}\left(\Delta d_{rob\_tar}(t)\right)\Delta^2 d_{rob\_tar}(t)\end{aligned} \tag{6}$$

Where β_3 and β_4 are weight coefficient, $0<\beta_3$, $\beta_4<1$.

③ $d_{rob\_obs}<d_{\min}$

The biggest value of punishment value will be set to $V=-1$ when distance between robot and obstacle is small than dangerous distance.

© 2013 ACADEMY PUBLISHER

Q function is designed as follow based on evaluate signal:

$$Q(s_i(t),a_k(t)) = \left(1-\gamma(p_{ik})\right)Q(s_i(t-1),a_k(t-1)) + \gamma(p_{ik})\left[V_t + \eta \max_{a_k} Q(s_i(t+1),a_k(t))\right] \quad (7)$$

η is the discounted factor which represent attentive degree of learning system to action. Learning system concerns the lately action when the value of discounted factor is small. Learning rate function $\gamma(p_{ik})$ controls the learning velocity of system. The value $\gamma(p_{ik})$ is bigger the convergence rate is faster. However, excessive value of $\gamma(p_{ik})$ will cause non-convergence. So learning rate function designs as:

$$\gamma(p_{ik}) = \frac{1}{1+\exp\left[\dfrac{p_{ik}(t)-p_{ik}(t+1)}{p_{ik}(t)}\right]} \quad (8)$$

The excitation rate p_{ik} of "state s_i -action a_k" pair is added to learning rate function $\gamma(p_{ik})$, which will make the Q value of each "state -action" pair has different adjusting rate. The adjusting process of Q value will embody tendency characteristic of animal operant conditioning. It can be seen from equation (8), the adjusting rate of Q value incline augment if the probability variable $p_{ik}(t)-p_{ik}(t+1)$ less than zero. Contrarily, the adjusting rate of Q value inclines decrease. Robot will learn quickly the optimal navigation strategy by adopting learning mechanism in (8).

E. Designation of Action Selected Strategy

Action selected strategy is the core part of robot navigation. The mail task of robot is exploring environment in the early period of learning, so randomness of action selecting is big; the mail task of robot is ensuring learning convergence in the later period of learning, so the randomness of action selecting is small. Boltzmann machine is adopted to realize annealing calculation. The selecting probability of action a_k is

$$p_{ik}(s,a) = \frac{e^{Q(s,a)/T}}{\sum_{a_k \in A} e^{Q(s,a_k)/T}} \qquad T = T_0 t^{-1/\varphi} \quad (9)$$

Where T is the temperature coefficient, T_0 is the initial temperature value, parameter φ is used to control the velocity of anneal. The temperature coefficient T determines the random degree of action selected. The value of T is bigger and the probability of each of action is closer, correspondingly the random degree of action selected is higher. The temperature T will decay from T_0 with the increasing of time t and decay to zero when $t \to \infty$, which show that the learning system has been changed from initial blindfold learning to deterministic learning.

F. Learning Procedure

In the process of autonomous navigation, robot use sonar sensor information to implement obstacle avoidance strategy when it approaches goal. The learning procedure is as follow.

Step1. Initialization

Including initial position of robot, iterative learning times t_f, sample time, weight coefficient β_1, β_2, β_3, β_4, initial temperature value T_0, temperature parameter φ, discounted factor η. Initial selected probability of each action $p_{ik}(0) \approx \dfrac{1}{r}$ $(i = 1,2\cdots n), (k = 1,2\cdots r)$.

Step2. Observe state

Observe current state based on sensor information: $S(t) = \left\{d_{rob\_obs\_l}, d_{rob\_obs\_f}, d_{rob\_obs\_r}, d_{rob\_tar}, \theta\right\}$, and compute information entropy of state.

Step3. Select action

According to Boltzmann distribution, select an action from action space A based on state $s(t)$ of time t.

Step4. State transition

State will transition after implementing action $a_k(t)$: $S(t) \times A(t) \to S(t+1)$, and information entropy of state $s(t+1)$ can be computed.

If leaning times exceed N, then delete the state whose information entropy keep maximum and jump **Step5**, otherwise directly jump **Step5**.

Step5. Update learning arithmetic

Firstly, judging the distance between robot and obstacles:

If $d_{rob\_obs} > d_{\max}$, the evaluate signal $V_{ik}(t+1)$ of "state $s_i(t)$ -action $a_k(t)$" will be calculated according to equation (4);

If $d_{\min} < d_{rob\_obs} < d_{\max}$, the evaluate signal $V_{ik}(t+1)$ of "state $s_i(t)$ -action $a_k(t)$" will be calculated according to equation (5);

If $d_{rob\_obs} < d_{\min}$, learning fail and set punishment value $V = -1$.

Finally, Update Q value of "state-action" pare according to formula (6).

Step6. Update probability value

Update probability value $p_{ik}(s_i(t+1),a_k(t+1))$ of "state-action" pare according to formula (8).

Step7. Judging state of robot

If robot knocks obstacle, keep the value of Q and jump to **Step2**. If robot doesn't arrive goal, jump to **Step3** and select a new action $a_{k'}(t+1)$ based on adjusted Q value and probability value $a_{k'}(t+1)$. Robot will repeatedly perform the procedure **Step3**-**Step7**. Otherwise jump to **Step8**.

Step8. Recursive transmit

If $\left|Q_{t+1}(s,a)-Q_t(s,a)\right| < \varepsilon$, jump to **Step9**; Otherwise jump to **Step2**, and begin a new leaning.

Step9. End.

G. Convergence Proof of Learning Algorithm

Theorem1: Suppose navigation strategy system is an operant conditioning probabilistic automaton, then:

© 2013 ACADEMY PUBLISHER

$$\lim_{t\to\infty} p_{ik}\left(s_i(t), a_k(t)\right) \to 1; \lim_{t\to\infty} p_{ik'}\left(s_i(t), a_{k'}(t)\right) \to 0 \quad (10)$$

Where action $a_k(t)$ increase Q function value and action $a_{k'}(t)$ decrease Q function value. The theorem1 show that navigation strategy system will select the action which will maximize Q function value with probability one and minimize Q function value with probability zero when $t \to \infty$.

Proof: Suppose the navigation strategy system $a_k \in A$ selecting action at state s_i, if $a(t) = a_k(t)$ and from (9), the following formula can be obtained

$$\lim_{t\to\infty} p_{ik}\left(s_i(t), a_k(t)\right) = \lim_{t\to\infty} \frac{e^{Q(s,a_k)/T}}{\sum_{a\in A} e^{Q(s,a)/T}} \quad (11)$$

The temperature T value will decay to zero when $t \to \infty$, so

$$\lim_{t\to\infty} p_{ik}\left(s_i(t), a_k(t)\right) \approx 1 \quad (12)$$

Because $\sum_{k=1}^{r} p_{ik}\left(s_i(t), a_k(t)\right) = 1$, so $\lim_{t\to\infty} p_{ik'}\left(s_i(t), a_{k'}(t)\right) \approx 0$.

Theorem2: Suppose navigation strategy system is an operant conditioning probabilistic automaton, then information entropy $H_i(\{A_i\} \mid s_i)$ will converge to minimum with time *t*.

$$\lim_{t\to\infty} H_i(t) = H_{i\min}$$

Proof: The navigation strategy system is a self-learning and self-organizing system based on Skinner OC theory. So process of self-organization is process of absorbing information and absorbing negative entropy. In other words, the purpose of self-organization is to eliminate uncertainty of system. Then there is necessary to prove the action entropy will converge to minimum for illustrating the self-organizing performance of navigation strategy system.

The information entropy value will reach the maximum when the probability value of each action has same value from the characteristic of entropy. The probability of action will change with learning, so the (1) becomes

$$H_i(t) = H_i\left(A(t) \mid s_i\right) = -\sum_{k=1}^{r} p_{ik}\left(s_i, a_k\right)\log_2 p_{ik}\left(s_i, a_k\right) = -\left[p_{ik}\left(s_i, a_k\right)\log_2 p_{ik}\left(s_i, a_k\right) + \sum_{k'=1, k'\neq k}^{r} p_{ik'}\left(s_i, a_{k'}\right)\log_2 p_{ik'}\left(s_i, a_{k'}\right)\right] \quad (13)$$

Applying result of Theorem1,

$$\lim_{t\to\infty} H_i(t) = \lim_{t\to\infty}\left[-p_{ik}\left(s_i, a_k\right)\log_2 p_{ik}\left(s_i, a_k\right) - \sum_{k'=1, k'\neq k}^{r} p_{ik'}\left(s_i, a_{k'}\right)\log_2 p_{ik'}\left(s_i, a_{k'}\right)\right] \to 0 \quad (14)$$

IV. Result Of Experiment And Analysis

In the platform MobotSim, mobile robot working environment was created for validating the feasibility of designed autonomous navigation strategy. The navigation environment is shown by grid map which consist of 70×50 grids, and the size of each grid is 0.2×0.2 m. There is static obstacle in the navigation environment. Robot is represented by a red little roundness with diameter is about and 0.5m, and the goal which set to be yellow is also a little roundness.

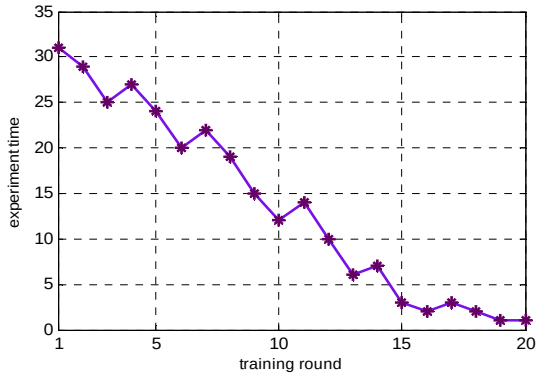
A. Simulation Parameters Setting

The initial iterative learning step is *t*=0; the sample time is t_s=0.1s; the weight coefficient are $\beta_1 = 0.65$, $\beta_2 = 0.36$, $\beta_3 = 0.72$ and $\beta_4 = 0.44$; the discounted factor is $\eta = 0.14$; learning threshold value is $\varepsilon = 0.01$; the state space of robot is $S = \{s_i \mid i = 1, 2, \cdots, 3125\}$; the action space is $A = \{a_k \mid k = 1, 2, \cdots, 7\}$; the learning round is *N*=5; the selected probability of each action is about $p_{ik}(0) \approx \frac{1}{7}$ and correspondingly the initial entropy value of each state is maximum with value $H_i(0) = -\sum_{k=1}^{5} p_{ik} \times \log_2 p_{ik}\Big|_{p_{ik}=\frac{1}{5}} \approx 2.72$.

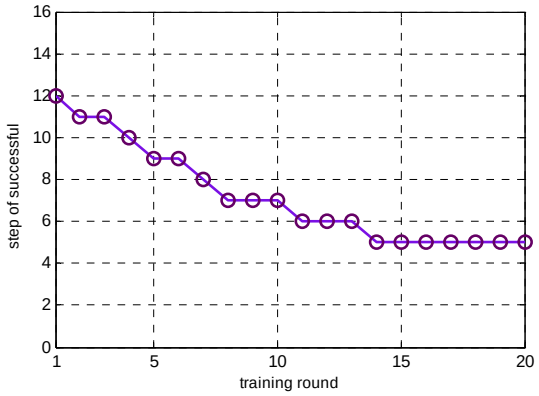
B. Simulation Results Analyses

Twenty round training were put up in experiment and one round training will end when robot arrives goal point or the learning times exceed one hundred. The next round training would begin based on the learned experience. Learning velocity would speed-up after ten round training, and training times of successful navigation need little training times after fifteen round training, and steps of successful navigation need little training rounds until tend to steady step (see Fig.5). The simulation results showed that robot learned select the action which could avoid colliding and arrived goal point with short path by autonomous learning.

© 2013 ACADEMY PUBLISHER



(a) Time of each round training



(b) Ending step of successful experiment

Figure 5. Many rounds navigation learning results.

The change of stated information entropy in the process of twenty round training sees Fig.6. Robot didn't have any prior knowledge in initial experiment. Information entropy value gradually decreased to minimum from maximum value and on longer changed after fifteen round training. The change of information entropy shows that the learning process of simulating operant conditioning was variational and dynamic. The self-organizing became higher with accumulation of learning experience.

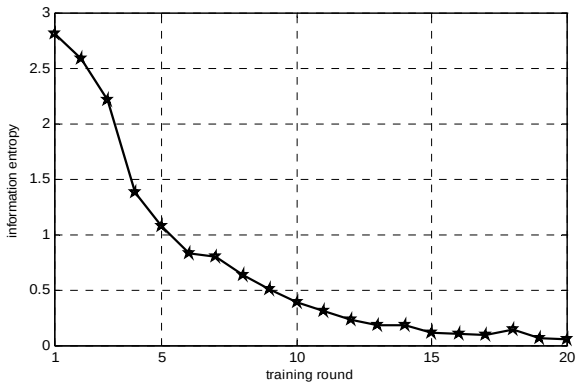
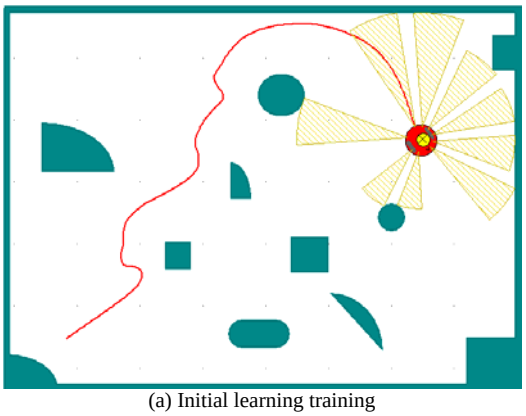
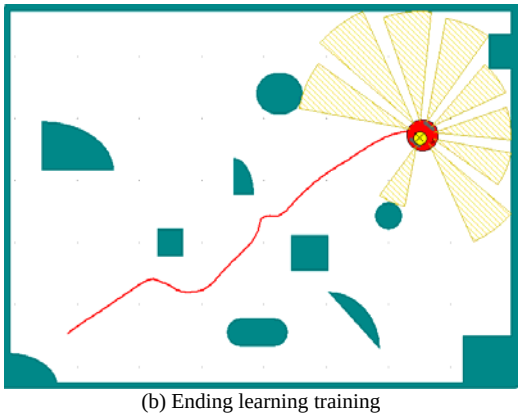


Figure 6. Many rounds navigation learning results.

Training result of initial and ending training phase was taken out for clearly showing the bionic learning process of self-learning and self-organization (see Fig.7). There are eleven irregular shape obstacles in simulation environment and the yellow sector part is the coverage area of sensor. The starting point of robot navigation is (10, 10) and the target point is (55, 35).



(a) Initial learning training



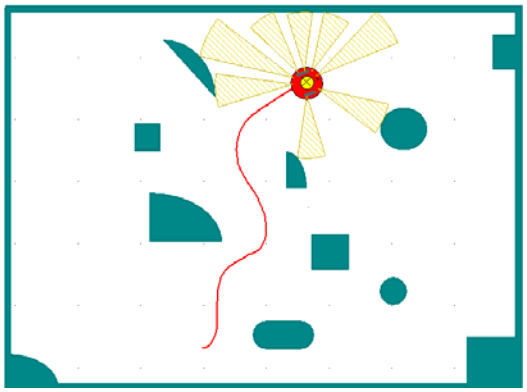
(b) Ending learning training

Figure 7. Navigation trajectory based on bionic autonomous learning strategy.

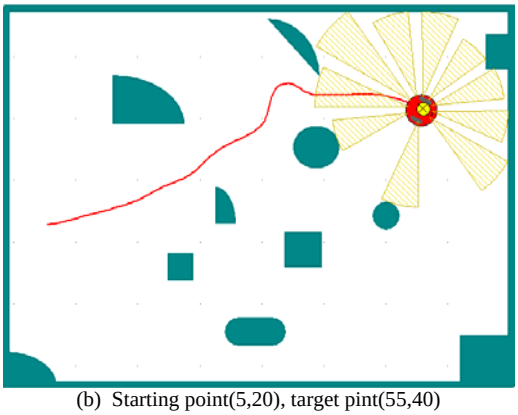
The simulation result of Fig.7(a) showed that the navigation trajectory wasn't at its optimal and occurs collision once a while in initial learning because robot hadn't any prior knowledge but had high performance of random. So robot could not select effective action to execute path planning. With the interaction between robot and environment, robot began incline to select the action which could obtain more reward and improve the performance of navigation. The simulation result of Fig.7 (b) showed that robot had could move to the target point and avoided static obstacle in the same time in terminal learning, which showed the effective action and collision free navigation path had been learned. The learning of robot was changed from blindfold random learning to deterministic learning and the operant conditioning formed. The learning process of robot autonomous navigation react the operant conditioning process of human or animal. The learning result showed that robot inclined to select the beneficial action through autonomous learning.

Save the learned "state-action" mapping relation and change the working environment: setting different starting point are (30,5) and (5,20), target point is (40,42); and (55,40) respectively. The distributions of obstacles are also changed, and the navigation experiment was conducted again. The simulation result sees Fig.8, which showed that robot could evade obstacles successfully and arrived at goal point in difference environment by using designed bionic learning strategy. The learning results

© 2013 ACADEMY PUBLISHER

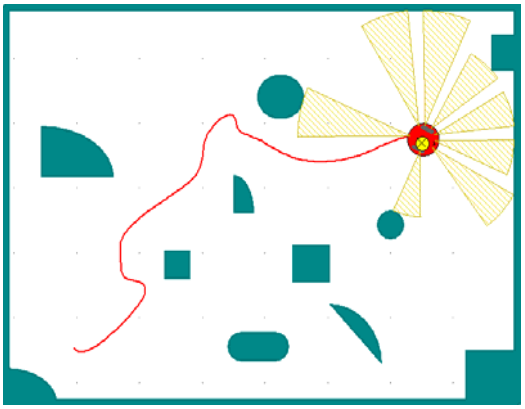


(a) Starting point (30,5), target pint (40,42)

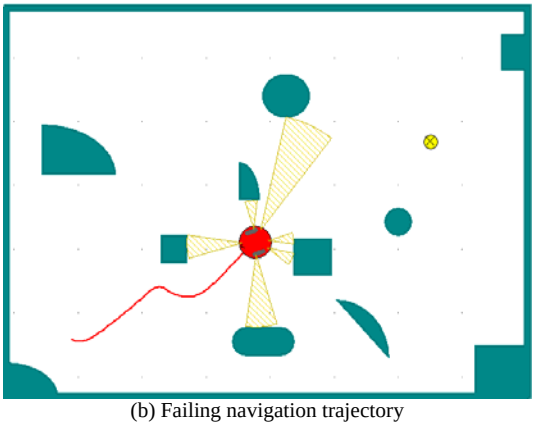


(b) Starting point(5,20), target pint(55,40)

Figure 8. Navigation trajectory after change the environment.



(a) Successful navigation trajectory



(b) Failing navigation trajectory

Figure 9. Navigation trajectory based on artificial field method.

established the validity of the designed navigation strategy.

In order to further verify the validity of autonomous navigation strategy, bionic leaning strategy and artificial potential field method [10] were compared in the same navigation environment as shown in Fig.7. The navigation trajectory based on artificial field method sees Fig.9. Comparison of navigation result Fig.7(b) and Fig.9(a) showed that the former trajectory was smooth and short while the later trajectory not only had swinging phenomenon but also was longer than former. The comparison further proved that bionic autonomous learning is an interaction process with environment and can learned the optimal short path with few swing. Otherwise, using artificial potential field method, robot inclined trapped in a point as shown in Fig.9(b) when it encountered dense obstacle. Robot will difficultly get rid of local minimum point and finally trapped in the point because of attract of goal point. However, the phenomenon will not happen in bionic learning strategy.

Above simulation results showed that navigation process based on bionic learning was a continuously explore and consolidate process. Robot could understand effectively the environment and finished reasonable mapping from state space to action space based on sufficient learning. So it was efficient and feasible use bionic autonomous learning solve autonomous navigation problem of robot in unknown environment.

IV. Conclusion

Bionic learning strategy which simulates operant conditioning mechanism is put forward to realize the autonomous navigation mobile robot in the paper. Robot keeps interact with environment in the process of navigation and achieve expected aim through repeated learning and training. The navigation simulation of unknown environment is executed in Mobotsim environment, and the simulation results prove that: 1) The designed bionic autonomous learning strategy has quick learning velocity and accurate navigation ability. 2) Robot can successfully evade obstacles and arrive at goal point with optimal path based on bionic autonomous learning strategy. 3) Robot can as much as possible ergodic environment by many rounds learning based on bionic autonomous learning strategy and the autonomous negative ability of robot is enhanced. In addition, simulation effect is better than artificial field method.

Acknowledgment

The work was supported by Special Fund of Fundamental Scientific Research Business Expense for Higher School of Central Government (Projects for creation teams) (No. ZY20110104), Teachers' Scientific

© 2013 ACADEMY PUBLISHER

Research Fund of China Earthquake Administration (No. 20110122).

REFERENCES

[1] S. Parasuraman, V. Ganapathy, and B. Shirinzadeh, "Multiple Sensors Data Integration Using MFAM for Mobile Robot Navigation," *IEEE Congress on Evolutionary Computation. Singapore*, vol. 2, pp. 2421-2427, September 2007.

[2] L.F. Zhou and J. Jiang, "An approach to safe path planning for mobile robot in the dynamic environment based on compact maps," *Journal of Computers*, vol.7, no.2, pp. 405-410, February 2002.

[3] C. Zeng, Q. Zhang, and X. P. Wei, "GA-based global path planning for mobile robot employing A* algorithm," *Journal of Computers*, vol.7, no.2, pp. 470-474, February 2002.

[4] Q.B. Zhu, "Algorithm for navigation of multi-robot movement in unknown environment," *Journal of Software*, vol.17, no.9, pp. 1890-1898, September 2006.

[5] M. L. Zhu, X. H. Zhang, X. Y. Wang, W. B. Tang, "Computer integration system for autonomous intelligent robot with self-organization structure," *Journal of Software*, vol.11, no.3, pp. 368-371, May 2000.

[6] M. Obayashi, N. Nakahara, T. Kuremoto, and et al, "A robust reinforcement learning using the concept of sliding mode control," *Artif Life Robotics*, vol. 13, no. 2, pp. 526-530, March 2009.

[7] D. Q. Zhu, T. Mei, L. Sun, "Fuzzy support vector machines control for 6-DOF parallel robot," *Journal of Computers*, vol.6, no.9, pp. 1926-1934, September 2011.

[8] Q. Zhang, M. Li, X. S. Wang, Y. Zhang, "Reinforcement learning in robot path optimization," *Journal of Computers*, vol.7, no.3, pp. 657-662, March 2012.

[9] M. M. Veloso, "CMRoboBits: Creating an Intelligent AIBO Robot," *AI Magazine*, vol. 27, no. 1, pp. 67-82, January 2006.

[10] B. Brembs, "The importance of being active," *Journal of Neurogenetics*, vol. 23, no. 1/2, pp. 120-126, February 2009.

[11] J. F. Qiao, R. Y. Fan, H. G. Han, X. G. Ruan, "Research and realization of dynamic neural network navigation algorithm for mobile robot," *Control Theory & Applications*, vol. 27, no. 1, pp. 111-115, January 2010.

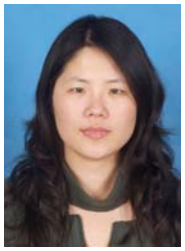
[12] B. Brembs and W. Plendl, "Double dissociation of PKC and AC manipulations on operant and classical learning in drosophila," *Current Biology*, vol. 18, no. 15, pp. 1168-1117, December 2008.

[13] K. Itoh and H. Miwa, M. Matsumoto, and et al, "Behavior Model of Humanoid Robots Based on Operant Conditioning," *IEEE/RAS Int Conf on Humanoid Robots. Piscataway*, vol. 20, pp. 220-225, December 2005.

[14] L. Dai, X. Ruan, and J. Chen, "Bionic Experiments Based on Autonomous Operant Conditioning Automata," *International Journal of Modelling, Identification and Control*, vol. 14, no. 4, pp. 286-293, May 2010.

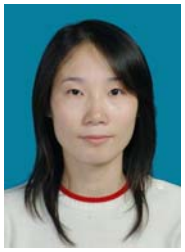
[15] X. Ruan and J. Chen, "Operant conditioning reflex learning control scheme based on SMC and Elman network," *Control and Decision*, vol. 26, no. 9, pp. 1398-1406, September 2011.

[16] X. Ruan, and J. Chen, "On-line NNAC for a Balancing Two-Wheeled Robot Using Feedback-Error-Learning on the Neurophysiological Mechanism," *Journal of Computers*, vol.6, no.3, pp. 489-496, March 2011.



Jianxian CAI was born in 1978. In 2010, she received her Ph.D degree in Pattern Recognition and Artificial Intelligence at Chinese Beijing University of Technology. She received her B.E. degree and M.E. degree in Control Theory and Control Engineering from Chinese HeBei University of Science & technology in 2001 and from Chinese Yanshan University in 2003, respectively. Her major fields of study are automatic control, robot intelligent control and machine learning.

She is a LECTURE of Department of Disaster Prevention Instrument at the Institute of Disaster Prevention which in China's Hebei Province. Her Published papers mainly have following paper: Bionic autonomous learning control of a two-wheeled self-balancing flexible robot (Guangzhou, China: Journal of Control Theory and Applications, 2011), Robust Bionic Learning System Design Based on FBFN and Its Application to Motion Balance Control (Shenyang, China: Robot, 2010), Bionic Autonomous Learning Mechanism Study based on Automaton and Applied on Robot (Switzerland, Europe: Applied Mechanics and Materials, 2011). Currently, his research interests are in the areas of machine learning and bionics.



Lixin LI was born in 1979. She received her B.E degree in Automation and M.E. degree in Measurement Technology and Automatic Instrumentation from Electric Engineering College of Yanshan University, in 2003 and 2006, respectively. Her major field of study was optical fiber sensing and detection.

She is a LECTURE of Department of Disaster Prevention Instrument at the Institute of Disaster Prevention in China's Hebei Province. Her Published papers mainly have following paper: Experiment and theoretical analysis of fiber Bragg grating under transverse force to a small grating section (Tianjin, China: Optoelectronics letters, 2005), Theoretical Analysis of Fiber Bragg Grating Characterization by Transverse Force to A Middle Section (Shanghai, China: Chinese Journal of lasers, 2005). Currently, his research interests are in the areas of machine learning and bionics.

© 2013 ACADEMY PUBLISHER