2003 International Symposium on Advanced Intelligent Systems
September 25-28, 2003, Jeju, Korea

Reinforcement learning control of a mobile robot in home network environment

Dong-Oh Kang and Jeunwoo Lee
Computer and Software Laboratory, Electronics and Telecommunications Research Institute
161 Gaejong-dong Yuseong-gu Daejon 305-350 Korea
email: dongoh@etri.re.kr, ljwoo@etri.re.kr

Abstract – **The following paper deals with a control problem of a mobile robot in home network environment. The home network causes the mobile robot to communicate with sensors to get the sensor measurements and to be adapted to the environment changes. We adopt the reinforcement learning scheme for the solution to the problem, and show some simulation results.**

I. INTRODUCTION

Recently, the technology for home network has made remarkable progress around the industries. And, every home is expected to be equipped with home network capability in recent future as seen in the concept of ubiquitous computing. Such rapid development of the technology for home network will make home smarter by equipping home with intelligent information devices, which includes intelligent service robots. Because service robots are aimed at helping humans in daily life, it is inevitable for service robots to operate in home network environment. Especially mobile robots will play an important role among the service robots [1].

The home environment renders various unstructured environment which may be changed by replacing sensor devices, network media, or changing the location of the sensors. The change in the environment deteriorates the performance of control of the mobile robot. The mobile robot in home network environment should cope with the problem. Because it is hard to model the changes in home network environment, we apply the reinforcement learning to the problem. Furthermore, the improvement in control performance will reduce the energy consumption of mobile robots, which is important for mobile robots because they use the battery as power source.

Reinforcement learning is how to learn the optimal policy based on the reward from environment. It needs no model for environment. Therefore, it is popularly applied to mobile robot control in unstructured environment [2]. In this paper, to get the improved performance of control of a mobile robot in spite of the change in home network environment, we use the fuzzy inference system with reinforcement learning [3]. Various situations are supposed as the change in home network environment, for example, sensor replacement, media change, and, home server replacement, etc. The change in home network environment is assumed as the change in time delay of sensor values and in noise level. To show the effectiveness of the proposed method, some simulation results are given, which are performed in real home network environment such as LAN, and wireless LAN, etc.

II. PRELIMINARIES

A. Reinforcement Learning

Reinforcement learning is how to learn the optimal policy based on the reward from environment [4]. Reinforcement learning is based on Markov decision process where the information available to the agent in the current situation is sufficient to determine the future states of the environment independent of the past information.

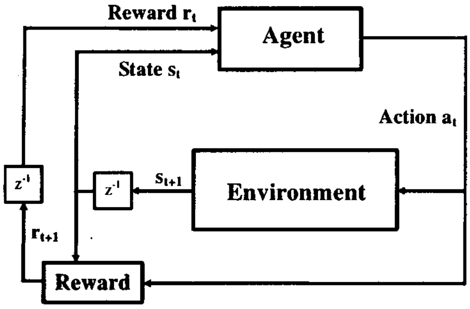


Figure 1. General structure of reinforcement learning.

A state-value function $V^{\pi}(s)$ of the policy π specifies what is good in the long run when the system starts at the state s and adopts the policy π.

In case of an infinite-horizon model, the expected discounted return is used for the state-value function as follow:

$$\begin{aligned} V^{\pi}(s) &= E_{\pi}(R_t \mid s_t = s) = E_{\pi}\left(\sum_{k=0}^{\infty}\gamma^k r_{t+k+1} \mid s_t = s\right) \\ &= E_{\pi}\left(\sum_{k=0}^{\infty}\gamma^k \Re_{t+1}(s_{t+k}, \pi(s_{t+k})) \mid s_t = s\right) \\ &= \sum_{a}\pi(s,a)\sum_{s'} P^{a}_{ss'}(\Re^{a}_{ss'} + \gamma V^{\pi}(s')), \end{aligned} \tag{1}$$

where s is an initial state, s_t is a state at time t after starting from the initial state s, r_t is a reward at time t given that the agent follows the policy π, where $P^{a}_{ss'}$ is the probability of transition from state s to state s' under action a, and $0 \le \gamma < 1$ is the discount rate.

If the policy π is optimal, it satisfies the following relation called Bellman optimality equation for V^*:

$$\begin{aligned} V^*(s) &= \max_{a \in A(s)} Q^*(s,a) \\ &= \max_{a}\sum_{s'} P^{a}_{ss'}(\Re^{a}_{ss'} + \gamma V^*(s')). \end{aligned} \tag{2}$$

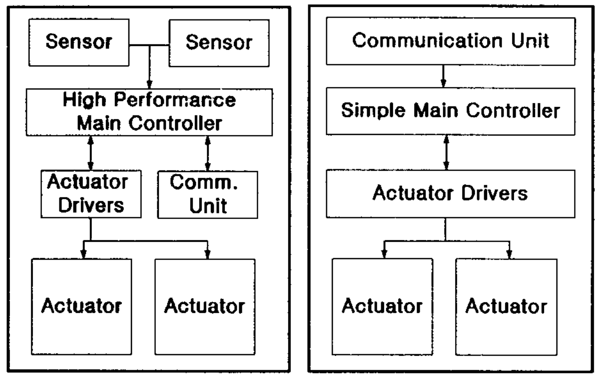
Therefore, reinforcement learning is how to get the optimal policy which gives the optimal value function. Adaptive heuristic critic and Q-learning are two major reinforcement learning methods [4].

In Q learning, we get the optimal policy using the Q value, i.e., the action-state value function, and learning rule as follows:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \max_{a} Q(s_{t+1}, a) - Q(s_t, a_t)] \tag{3}$$

B. Mobile robots in home network environment

In general, a mobile robot has some sensors to detect its location and posture, whereas sensors are equipped within home network in home network environment. Therefore, a mobile robot communicates with sensors through home network. A mobile robot in home network environment has three parts: main controller, actuators, and communication units, while the conventional robot has four parts: main controller, actuators, sensors, and communication units [1].



(a) (b)

(a) General mobile robot architecture.

(b) Mobile robot architecture in home network.

Figure 2. Sensing architecture of mobile robots in the home network environment [1].

Although the mobile robot in home network environment has no sensors such as gyroscopes and laser range finders which are expensive and give limited information about environment, the robot can get the global information through home network, which is provided by some devices such as the home server. The home server provides the global map by gathering sensor information from various sensors pervaded in home through home network, and the middleware can give interoperability among heterogeneous devices.

III. FUZZY CONTROLLER WITH REINFORCEMENT LEARNING

For the fuzzy controller to be designed, the fuzzy inference system is used in the paper, which has multiple consequent singletons for the consequent fuzzy set [3]. Among the multiple consequent singletons, one singleton is selected as the consequent fuzzy set for the fuzzy rule. The rule is in the form of the following:

R_i : If x_1 is L_1^i and ...and x_N is L_N^i, then u is U^i

$$U^i \in \{U^{i,1}, U^{i,2} \cdots U^{i,p}\} \quad (4)$$

At any instance, only one output consequent term should be selected in the scheme.

$$y = \frac{\sum_{i=1}^{N} U^i \mu_i(\mathbf{x})}{\sum_{i=1}^{N} \mu_i(\mathbf{x})} \quad (5)$$

where U^i is the selected output consequent term of the rule i among $\{U^{i,1}, U^{i,2} \cdots U^{i,p}\}$.

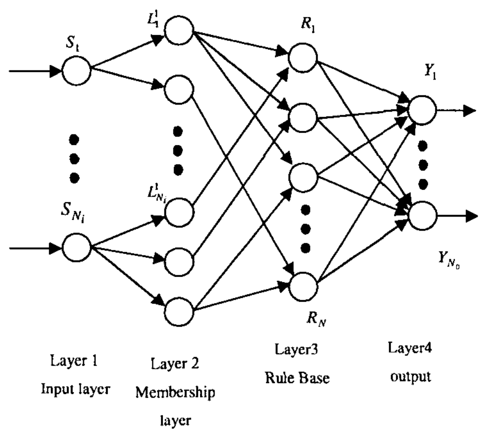


Figure 3. Adaptive Fuzzy Inference System [3], where N_i: number of the input variables, N: number of the rules, N_o: number of the outputs.

To optimize the fuzzy inference system, fuzzy Q-learning algorithm is adopted in this paper [3]. The temporal difference for the adaptation of the fuzzy inference system is calculated as follows:

$$\tilde{\varepsilon}_{t+1} = r_{t+1} + \gamma Q_t^*(s_{t+1}) - Q_t(s_t, U_t) \quad (6)$$

IV. SIMULATION

A. Mobile robot

For simulation, the kinematics of a mobile robot is used as (7). Input variables are the velocity values of both wheels. We assume there is a kind of damping factor when we drive the wheels, therefore, the first order dynamic equation as (8) is used.

$$\dot{P} = \begin{bmatrix} \dot{x}_c \\ \dot{y}_c \\ \dot{\theta}_c \end{bmatrix} = \begin{bmatrix} \cos\theta_c & -h\sin\theta_c \\ \sin\theta_c & h\cos\theta_c \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} \quad (7)$$

$$\begin{bmatrix} v \\ w \end{bmatrix} = \begin{bmatrix} (v_R + v_L)/2 \\ (v_R - v_L)/L \end{bmatrix}$$

$$\begin{aligned} \dot{v}_R &= -0.8 v_R + 0.8 u_R \\ \dot{v}_L &= -0.8 v_L + 0.8 u_L \end{aligned} \quad (8)$$

where v_R is the right wheel velocity value, v_L is the left wheel velocity value, h is the displacement between the center of the robot and the wheel axis, and L is the distance between two wheels. For simulation, h=0 (m), L=0.3 (m). The reward for reinforcement learning is given as follows:

$$r = 1. - e/0.7 \quad (9)$$

where e is the distance between the desired path and the center of the mobile robot.

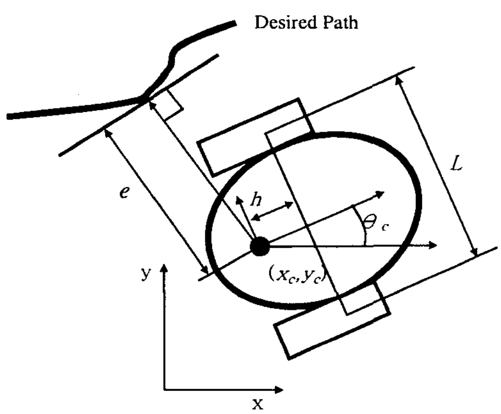


Figure 4. Kinematics of a mobile robot.

In simulation, we use two personal computers. One personal computer emulates a mobile robot, and the other emulates the sensors. Two computers communicate through home network just like a mobile robot and sensors in home network. Fuzzy controller uses the data from remote computer which emulate sensors in home network, whereas the data are originated from the computer where the fuzzy controller and mobile robot kinematics are emulated.

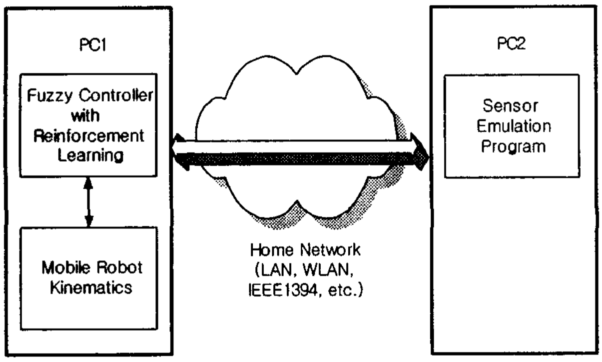


Figure 5. Simulation Environment.

B. Simulation results

We assume the sensor noises have Gaussian random distribution with zero mean and standard deviation (0.01, 0.02, 0.03) for the posture variable (x, y, θ). Initial posture of the mobile robot is set as (0,-0.4,0) and the desired path is set to the x-axis. Therefore, the distance between the desired path and the center of the mobile robot, that is, e is the same as $|y|$. The trials are performed for 10 times where each trial is composed of ten learning periods. The average squared error sum among trials is used as the performance index.

Table 1. Simulation results

Simulation type	Performance Index
Random delay with no learning	75.36
Random delay with learning	50.83
LAN with no learning	58.167
LAN with learning	55.4
WLAN with no learning	69.1
WLAN with learning	64.3

Table 1 shows the simulation results. The first row is the simulation with random delay 0.2~0.7 (sec.) and no learning performed. The second row is the simulation with reinforcement learning. The performance is improved by 33% reduction of the average squared error sum. In real LAN environment, the value in the third column is derived without learning. With reinforcement learning, the performance is improved as the value in the fourth column. In WLAN environment, the values in the fifth and sixth columns show the improvement in the performance.

V. CONCLUSION

In this paper, the reinforcement learning scheme is proposed as a solution to the control problem of a mobile robot in home network environment. Some simulation results are given to show the effectiveness of the proposed scheme. The experiment with a real mobile robot and convergence issue of the reinforcement learning remain for the future research.

REFERENCES

[1] Byoung-Ju Lee, Hyun-Gu Lee, Joo-Ho Lee and Gwi-Tae Park "New architecture for mobile robots in home network environment using Jini," Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation, 2001, pp.471–pp.476, vol.1, 2001.

[2] H. R. Beom and H. S. Cho, "A Sensor-Based Navigation for a Mobile Robot Using Fuzzy Logic and Reinforcement Learning," IEEE Transactions on Systems, Man, and Cybernetics 25(3): pp.464-pp.477, 1995.

[3] L. Jouffe, "Fuzzy Inference System Learning by Reinforcement Methods," IEEE Transaction on Systems, Man, and Cybernetics, Part C, vol. 28, no. 3, pp. 338 – 355, 1998.

[4] R. S. Sutton and A. G. Barto, *Reinforcement Learning An Introduction*, MIT Press, 1998.