

A Study on Multi-Dimensional Fuzzy Q-learning for Intelligent Robots

Kazuo Kiguchi, Hui He, and Kenbu Teramoto

Abstract

Reinforcement learning is one of the most important learning methods for intelligent robots working in unknown/uncertain environments. Multi-dimensional fuzzy Q-learning, an extension of the Q-learning method, has been proposed in this study. The proposed method has been applied for an intelligent robot working in a dynamic environment. The rewards from the evaluation functions and the fuzzy Q-values generated by the neural networks (fuzzy Q-net) are expressed in vector forms in order to obtain optimal behaviors for several different purposes. By applying this learning method, evaluation and learning of fuzzy Q-values for the other behaviors can be carried out simultaneously in one trial. We express fuzzy states as the vector of fuzzy sets for input variables of the fuzzy Q-net. The behavior selection algorithm is also proposed in this study. The simulation results show the effectiveness of the proposed methods for a mobile robot selects optimal behavior in a dynamic environment.

Key words: Intelligent robot, Fuzzy Q-learning, Multi-dimensional learning, Behavior selection.

1. Introduction

Nowadays, intelligent robots are applied in many fields. The intelligent robots have many potential applications in industry, medicine, and even service at home that make their study important. However, highly intelligent tasks are still difficult to be achieved by the robots. When the robots perform the tasks in an uncertain/unknown environment, searching the optimal behavior is very important. It is not easy to find the optimal behavior in the changing environment under various situations. Therefore, some forms of learning are required in many applications. There are several kinds of learning methods for the robots, such as supervised learning, unsupervised learning, and reinforcement learning [1]. The supervised learning learns from exam-

ples provided by a knowledgeable external supervisor, the unsupervised learning learns without external supervisor, and the reinforcement learning learns from the evaluated feedback information called the reinforcement signal (critic). In order to find the optimal behavior practically, reinforcement learning is the most suitable method. Reinforcement learning is studied in most current research in machine learning, statistical pattern recognition, and artificial neural networks [1]. In many studies [1]-[20], a lot of methods of reinforcement learning such as multi-layered reinforcement learning, fuzzy reinforcement learning, Q-learning have been proposed for the intelligent robots to search optimal behavior. Multi-dimensional Q-learning has been applied to intelligent robots in order to obtain the optimal behavior in the static working environment in [6]. Since the distances between the robot and the obstacles are not discrete values practically, the multi-dimensional fuzzy Q-learning is applied to intelligent robots in this study, and fuzzy linguistic variables are prepared for the distance between the robot and the target or the obstacles.

Intelligent robots are expected to work intelligently even in the unknown/uncertain environment. It is sometimes difficult to evaluate their performance with only one evaluation function under various situations. The desired behavior sometimes depends on the circumstance while there can be contradicting objectives that have special importance in certain circumstance. For example, the behavior of less energy consumption is usually preferred. However, time efficiency is more important than energy efficiency when the robot is in a rush. Usually the best behavior with respect to energy consumption is not the same as that with respect to time efficiency. Furthermore, safety is the most important when the robot carries out important tasks even if the robot consumes more energy. Thus the desired behavior should be changed according to the situation. For this reason, in this paper, rewards from an evaluation function are expressed in a vector form [6] in order realize multi-dimensional fuzzy Q-learning.

In this paper, a mobile robot [8], [9], [21] have been used as an example of the intelligent robot. A key element in the reinforcement learning is an evaluation function. The purpose of this function is to measure the long-term utility or value of any given state. It is important because the robots use it to learn what to do next. The fuzzy evaluation function is used in the proposed

Corresponding Author: Kazuo Kiguchi is with the Department of Advanced Systems Control Engineering, Saga University, 1 Honjomachi, Saga 840-8502, Japan.

E-mail: kiguchi@me.saga-u.ac.jp

Manuscript received 2006; revised 14 May, 2007; accepted 2007.

fuzzy Q-learning algorithm. The robots can incrementally learn the optimal behavior based on the evaluation function by continual exercise. Fuzzy Q-learning for intelligent mobile robots working in the dynamic environment is described in this paper. The fuzzy Q-value derived by the special neural network (fuzzy Q-net) is used for the robot to select an optimal behavior. In the fuzzy Q-net, the weights and the output are represented by vector forms. Each component of vectors is in the charge of each item of the evaluation function [6]. Consequently, each component of the weight vectors and the output vector of the fuzzy Q-net is adjusted based on reward or punishment for each item of the evaluation after a certain behavior is performed. Therefore, the learning occurs in all vector component of the fuzzy Q-net while implementing any given behavior. This kind of cross-criticism and parallel learning make the multi-dimensional learning process more efficient. In this paper, three kinds of evaluation function are prepared for the energy conscious behavior, the hasty behavior, and the safety conscious behavior. In the fuzzy Q-net, input variables have been changed to fuzzy values [7].

In this paper, a fuzzy behavior selection algorithm is also proposed for the intelligent robots to select a suitable behavior in accordance with the situation in the dynamic working environment. Simulation has been performed to evaluate the effectiveness of the proposed methods. In the simulation, the mobile robot is supposed to head toward the goal subjected to various performance criteria. Some moving obstacles are prepared in the working environment. The hasty behavior, energy minimum behavior, and safe behavior are considered in the simulation.

This paper is organized as follows. Section 2 covers a dynamic model of mobile robot. In section 3, we describe the reinforcement learning, and define the evaluation function (reinforcement function). Fuzzy Q-net architecture and fuzzy Q-learning algorithm are described. In fuzzy Q-net, input variables are represented by fuzzy values. In section 4, the fuzzy behavior algorithm is explained. In section 5, in order to evaluate the effectiveness of proposed learning method, computer simulation has been performed, the simulation results are explained. We make conclusions in section 6.

2. Dynamic Model of Mobile Robot

A mobile robot [6], [8], [10] is applied as an intelligent robot in this study. The schematic diagram of the mobile robot is shown in Fig. 1, where v_l and v_r are the velocity of the left and right wheel of the mobile

robot, respectively. ϕ is the azimuth of the mobile robot, I_v is the moment of inertia around of robot, and b is the distance between the left wheel and right wheel of the mobile robot.

Let $x(t) = [v(t), w(t)]^T$ (1)

be the state variable vector and

$$u(t) = [u_r, u_l]^T \quad (2)$$

Be the manipulated variable vector. Where $v(t)$ is the linear velocity of robot, $w(t) = \dot{\phi}(t)$ is the angular velocity of robot, of the robot. Then the state space model for the mobile robot can be written as:

$$\dot{x}(t) = A * x(t) + B * u(t) \quad (3)$$

with

$$A = \begin{bmatrix} -2c/(mr^2 + 2I_v) & 0 \\ 0 & -cb^2/(2I_v r^2 + I_v b^2) \end{bmatrix}$$

$$B = \begin{bmatrix} -r/(mr^2 + 2I_v) & -r/(mr^2 + 2I_v) \\ -rb/(2I_v r^2 + I_v b^2) & rb/(2I_v r^2 + I_v b^2) \end{bmatrix}$$

where m is the mass of the robot, c is the viscous friction coefficient, r is the radius of the wheel, u_r and u_l are the right and left driving input torque, respectively.

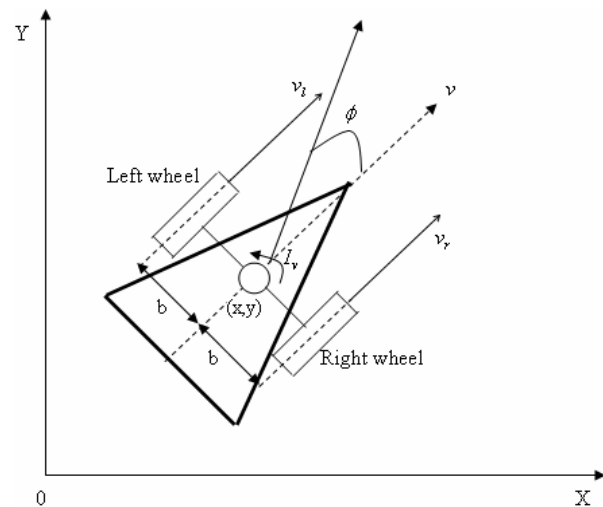


Fig.1 Schematic diagram of the mobile robot

The physical parameters of the mobile robot used in this study are given by $I_v = 0.654 [kgm^2]$, $m = 25.5 [kg]$, $b = 0.165 [m]$, $r = 0.05 [m]$, $I_w = 0.442 * 10^{-3} [kgm^2]$, $c = 0.048 [kgm^2/s]$.

3. Reinforcement Learning

Fuzzy Q-learning method, an extension of the Q-learning method that is one of the basic reinforcement learning methods, has been applied in this study.

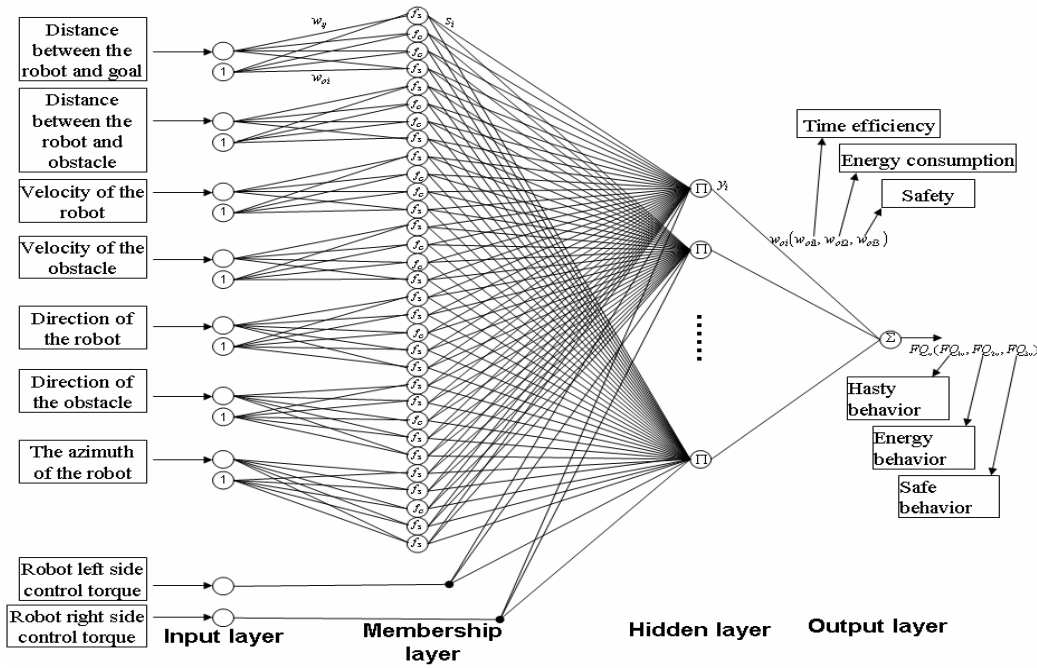


Fig. 2 Fuzzy Q-net architecture

A special neural network (fuzzy Q-net) is used to realize critic networks in this paper. In the fuzzy Q-net, input variables are represented by fuzzy values, and the weights and output are represented by vectors in which components are scalars. Each component of vectors is in charge of each item of the evaluation (reward). In this study, evaluation is carried out with respect to time efficiency (hasty behavior), energy consumption (energy minimum behavior), and safety (safe behavior) assuming that there are some moving obstacles in the working environment of the mobile robot. In this case, each weight vector and output vector of the fuzzy Q-net consist of three components (1, component: for time efficiency, 2, component: for energy consumption, and 3, component: for safety). After a certain behavior is performed, each component of the weight vectors and the output vector of the fuzzy Q-net is adjusted based on reward or punishment for energy minimum behavior, hasty behavior, and safe behavior.

3.1 Fuzzy Q-net architecture

The proposed fuzzy Q-net consists of four layers (input layer, membership layer, hidden layer, and output layer) is shown in Fig. 2. There are 7 input variables (1: distance between the robot and goal, 2: distance between the robot and the obstacle, 3: velocity of the robot, 4: velocity of the obstacles, 5: direction of the robot, 6: direction of the obstacle, 7: the azimuth of the robot). The input variables are represented by fuzzy values from fuzzy states. In the fuzzy Q-net architecture, Σ means sum of the inputs, Π means multiplication of the

inputs. Two kinds of the nonlinear function (f_G and f_S) are applied to express the membership function. The weight of the hidden layer is the vector. Where $w_{oi1}, w_{oi2}, w_{oi3}$ are the components of the vectors that represent time efficiency, energy consumption and safety, respectively. The output of the fuzzy Q-net is represented by the vector form where $FQ_{1v}, FQ_{2v}, FQ_{3v}$ are the components of the vector that represent hasty behavior, energy conscious behavior and safe behavior, respectively.

The fuzzy values were represented by linguistic variables. For example, “very small distance”, “small distance”, “big distance”, and “very big distance” are prepared for the distance between the robot and goal. Abbreviation of the linguistic variables, such as “VSD”, “SD”, “BD”, and “VBD”, are used as a fuzzy set. Then the fuzzy state 1 is:

$$\tilde{s}_1 = (VSD, SD, BD, VBD) \quad (4)$$

Similarly, we can define the other fuzzy states as a vector of fuzzy sets $\tilde{s}_2, \tilde{s}_3, \dots, \tilde{s}_n$, the fuzzy state from the real world environments is

$$\tilde{s} = (\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_n) \quad n = 1, \dots, 7 \quad (5)$$

Fuzzy sets for the distance between the robot and the goal and the distance between the robot and the obstacle are shown in the Fig. 3. Fuzzy sets for the velocity of the robot and velocity of the obstacles are shown in the Fig. 4. Here VSV means very small velocity, SV means small velocity, MV means middle velocity and BV means big velocity. Fuzzy sets for the azimuth of the

robot, direction of the robot, and direction of obstacles are shown in the Fig. 5. Here NS means negative small, NB means negative big, ZO means zero, PS means positive small and PB means positive big. Then, the robot can perceive the dynamic environment by using the fuzzy states.

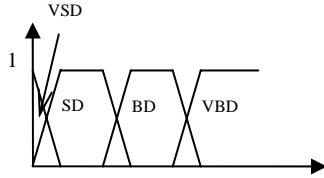


Fig. 3 Membership functions for the distance

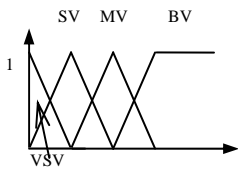


Fig. 4 Membership functions for the velocity

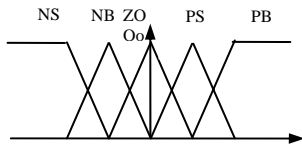


Fig. 5 Membership functions for the azimuth and direction

There are 50 neurons in the hidden layer of the fuzzy Q-net. The activation function used in the neurons is written as:

$$y_i = \frac{1}{1 + e^{-s_i}}, \quad i = 1, \dots, 50 \quad (6)$$

$$s_i = w_{oi} + \sum_{j=1}^{33} w_{ij} x_j \quad (7)$$

where w_{oi} is the bias weight vector of the i_{th} activation function, w_{ij} represents the connecting weight vectors between the i_{th} activation function and the j_{th} input given by x_j

The output of the fuzzy Q-net is the fuzzy Q values for the possible combination of the control input. The fuzzy Q values are calculated by:

$$FQ_v = \sum_{i=1}^{50} w_{oi} y_i \quad (8)$$

where w_{oi} is the output weight vectors of the fuzzy Q-net that connect the activation function and the output node.

The right and left side control torque inputs to the mobile robot obtained by a conventional controller based on the potential field method are denoted by u_{qr} and u_{ql} , respectively. The right side and left side control torque inputs given by the fuzzy Q-net are described by $u_{qr} \in U_r$ and $u_{ql} \in U_l$, respectively, where U_r and U_l are real bounded spaces in which the right and left torques are defined.

Input to the right and left torque of mobile robot wheels are given by:

$$u_l = u_{pl} + u_{ql} \quad (9)$$

$$u_r = u_{pr} + u_{qr} \quad (10)$$

The output of the fuzzy Q-net for a given vector of environmental information and a chosen control input is denoted by:

$$FQ_v(t) = [FQ_{1v}(t), FQ_{2v}(t), FQ_{3v}(t)]^T \quad (11)$$

The maximum $FQ_v(t)$ that is obtained by changing the right and left wheel torques in U_r and U_l for a given environment situation is denoted by $FQ_{v,max}(t)$. In this study, the output vector consists of three component functions (1: Hasty behavior, 2: Energy conscious behavior, 3: Safety conscious behavior) as shown in Fig. 6.

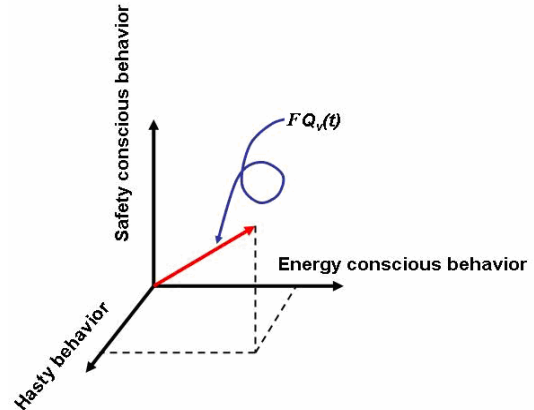


Fig. 6 The output vector from the fuzzy Q-net

3.2 Reinforcement function

The reinforcement function (evaluation function) determines the policy learnt by the behaviors. In this study, we use a vector of the reinforcement functions consists of three component functions (1: Hasty behavior, 2: Energy conscious behavior, 3: Safety conscious behavior) [6]. According to the preliminary study and experience, the following reinforcement functions are defined. The vector of reinforcement functions is given by:

$$r(t) = [r_1(t), r_2(t), r_3(t)]^T \quad (12)$$

Each component is given as:

$$r_1(t) = 5(v_{tar} + e^{-D}) + r_{obs} \quad (13)$$

for the hasty behavior, where v_{tar} is the velocity to the target, and r_{obs} is the reward or penalty for avoiding or colliding with the obstacle, which is calculated by $r_{obs} = -100 e^{-4|d_{obs}-0.4|}$ if close to the obstacle, and $r_{obs}=1$ if sufficient distance is kept, where d_{obs} is the distance to the obstacle.

$$r_2(t) = \frac{5}{1 + 100 e^{(|u_r|+|u_l|)}} + r_{obs} + e^{-D} + P \quad (14)$$

for the energy conscious behavior, where D is the distance to the target, P is a punishment given by $P = -10$ if $(|u_r| > 0.01$ or $|u_l| > 0.01)$.

$$r_3(t) = -100e^{-4|d_{obs}-0.4|} + r_{obs} + e^{-D} \quad (15)$$

for the safety conscious behavior.

The vector $r(t)$ is used to get enough fuzzy reward values for a certain motion. The linguistic variable for fuzzy reward value is defined to transferred into proper fuzzy sets, NV (negative value), Z0 (zero), PS (positive small), PM (positive middle), and PB (positive big). The fuzzy sets for the fuzzy reward ($Fr(t) = [Fr_1(t), Fr_2(t), Fr_3(t)]^T$), are shown in the Fig. 7.

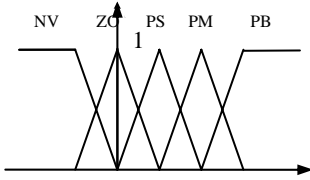


Fig. 7 Fuzzy reward

3.3 Fuzzy Q-learning algorithm

The algorithm of fuzzy Q-learning is explained in Table 1.

Table 1 Fuzzy Q-learning algorithms

1. For each fuzzy state \tilde{s} and action a , Initialize fuzzy $Q(\tilde{s}, a)$ arbitrarily
 2. Observe the current fuzzy state \tilde{s}
 3. Repeat:
 - a. Select an action a and execute it
 - b. Receive immediate fuzzy reward $Fr(\tilde{s}, a)$
 - c. Observe the new fuzzy state \tilde{s}'
 - d. Update the Fuzzy $Q(\tilde{s}, a)$ as the following equation
 - e. $FQ(\tilde{s}_t, a_t) = FQ(\tilde{s}_t, a_t) + \alpha[Fr_t + \gamma \max_{a_{t+1}} FQ(\tilde{s}_{t+1}, a_{t+1}) - FQ(\tilde{s}_t, a_t)]$
- α : learning rate [0 1], γ : discount rate [0 1]

4. Behavior Selection Algorithm

The proposed behavior selection algorithm is explained in this section. The fuzzy Q-net is supposed to output multiple fuzzy Q-values. Every fuzzy Q-value corresponds to the certain behavior. For example, the fuzzy Q_1-value corresponds to hasty behavior, the fuzzy Q_2-value corresponds to energy minimum behavior, and fuzzy Q_3-value corresponds to safe behavior. The optimal behavior (i.e., the behavior that generates the biggest fuzzy Q-value by the fuzzy Q-net) is obtained for every evaluated behavior (the hasty behavior, the energy minimum behavior, and the safe behavior). In the proposed behavior selection algorithm, the robot selects the optimal behavior of the selected behavior in accordance with the situation. For example, when the robot meets quickly approaching obstacles, the robot selects the hasty behavior to avoid them. When the obstacles passed, the robot selects the energy minimum behavior. When the robot performs important tasks, the robot selects the safe behavior.

The fuzzy behavior selection rules are expressed as follow. However, there are 103 fuzzy behavior selection rules. Some the representative fuzzy behavior selection rules of them are expressed in this paper.

If DRG is VB and DRO is VB and ... and RV is small then hasty behavior is selected.

If DRG is VB and DRO is VS and ... and RV is big then safe behavior is selected.

If DRG is VS and DRO is VS and ... and RV is middle then energy minimum behavior is selected.
where: DRG is the distance between the robot and goal.

DRO is the distance between the robot and obstacle.

RV is the velocity of the robot. VB is very big, VS is very small.

5. Simulation

In order to evaluate the effectiveness of the proposed methods, computer simulation has been performed. In this simulation, the mobile robot is supposed to head toward the goal subjected to various performance criteria. In order to evaluate effectiveness of the proposed method, dynamic working environment, in which 5 obstacles are moving as shown in Fig. 8, has been prepared, and the Q-learning is used as ordinary method to compare in the same working environment. The hasty behavior, the energy minimum behavior, and the safe behavior are considered in the simulation.

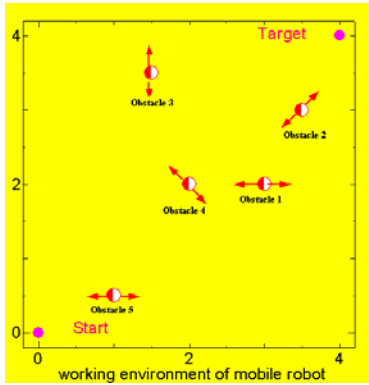


Fig. 8 Working environment

The moving obstacles are moving linearly over time in the working environment. The equations of the moving obstacles in the working environment are given by:

$$\begin{cases} s_{1x}(t) = s_{1x}(t) \pm v_{1x}(t) \times T \\ s_{1y}(t) = s_{1y}(t) \pm v_{1y}(t) \times T \\ s_{2x}(t) = s_{2x}(t) \pm v_{2x}(t) \times T \\ s_{2y}(t) = s_{2y}(t) \pm v_{2y}(t) \times T \\ s_{3x}(t) = s_{3x}(t) \pm v_{3x}(t) \times T \\ s_{3y}(t) = s_{3y}(t) \pm v_{3y}(t) \times T \\ s_{4x}(t) = s_{4x}(t) \pm v_{4x}(t) \times T \\ s_{4y}(t) = s_{4y}(t) \pm v_{4y}(t) \times T \\ s_{5x}(t) = s_{5x}(t) \pm v_{5x}(t) \times T \\ s_{5y}(t) = s_{5y}(t) \pm v_{5y}(t) \times T \end{cases} \quad (16)$$

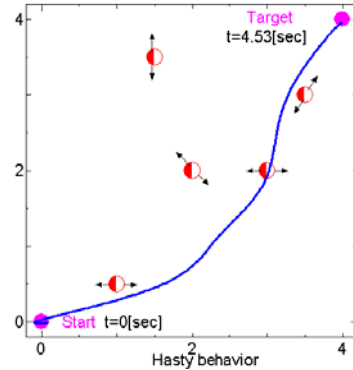
where $v_{1x} = 0.4 \text{ m/s}$, $v_{1y} = 0.0 \text{ m/s}$, $v_{2x} = 0.2 \text{ m/s}$, $v_{2y} = 0.2 \text{ m/s}$, $v_{3x} = 0.0 \text{ m/s}$, $v_{3y} = 0.3 \text{ m/s}$, $v_{4x} = -0.2 \text{ m/s}$, $v_{4y} = -0.2 \text{ m/s}$, $v_{5x} = -0.3 \text{ m/s}$, $v_{5y} = 0.0 \text{ m/s}$, T is the sampling time, $s_{ix}(t)$ is the position of x axis, and $s_{iy}(t)$ is the position of y axis.

In order to prove the effectiveness of the proposed fuzzy Q-learning method, the first simulation is carried out without the behavior selection algorithm to compare with the ordinary Q-learning method. The obtained behavior and the torque profiles with the Q-learning and fuzzy Q-learning in the simulation are shown in Fig. 9, Fig. 10, Fig. 11, and Fig.12, respectively.

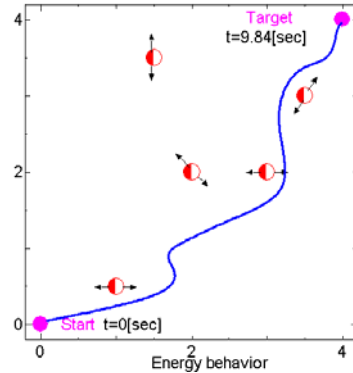
In the second simulation, the mobile robot selects the optimal behavior using proposed behavior selection algorithm observing the given state in the working environment. The simulation results are shown in Fig. 13.

Time efficiency is more important than energy efficiency when the robot is in a rush. In hasty behavior, the robot arrives at target quickly although there are moving obstacles in working environment. In energy minimum behavior, saving energy is more important for the robot. When the robot meets the approaching obstacles, the robot waits the obstacle until the obstacle

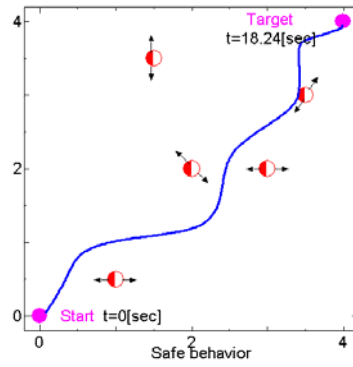
passed then goes to the target. In the safe behavior, the robot tries to stay away from the moving obstacles.



(a) Hasty behavior



(b) Energy behavior



(c) Safe behavior

Fig. 9 Simulation results with the Q-learning

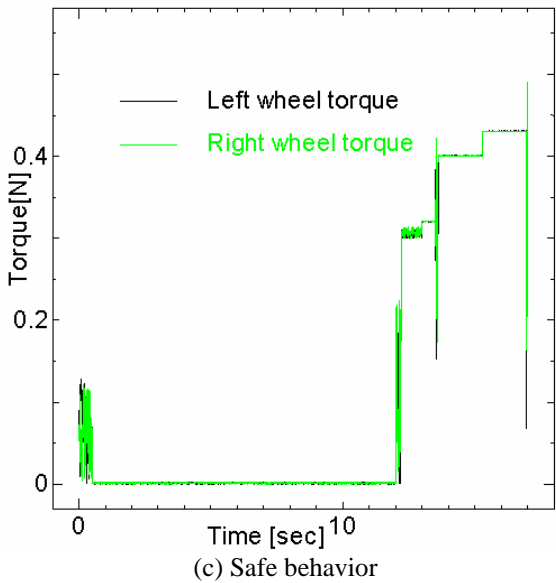
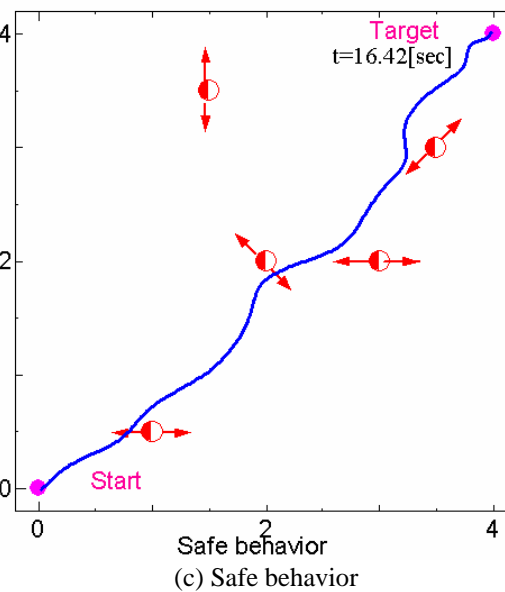
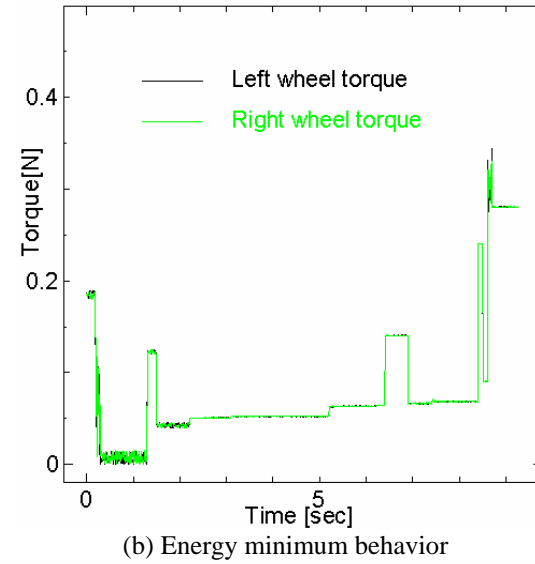
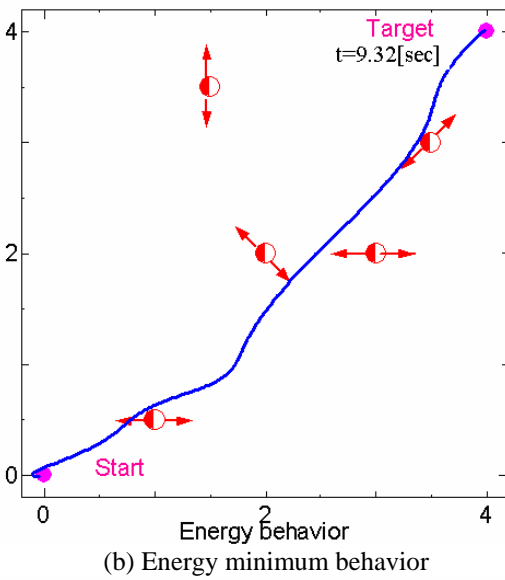
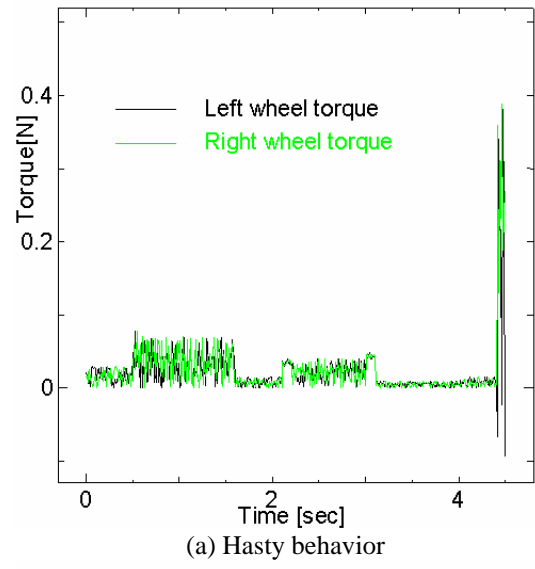
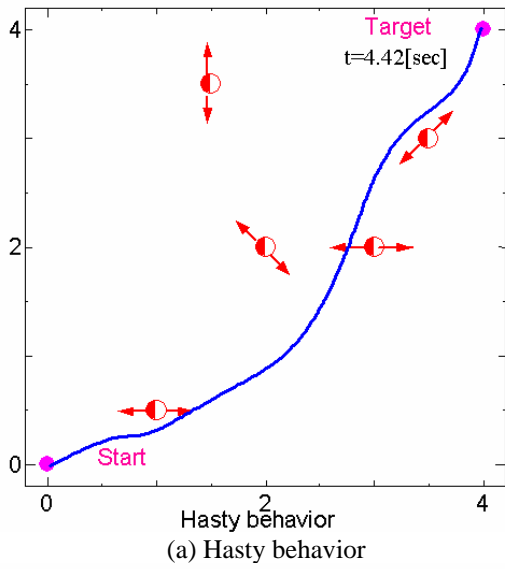


Fig. 10 Simulation results with fuzzy Q-learning

Fig. 11 Torque profiles with Q-learning

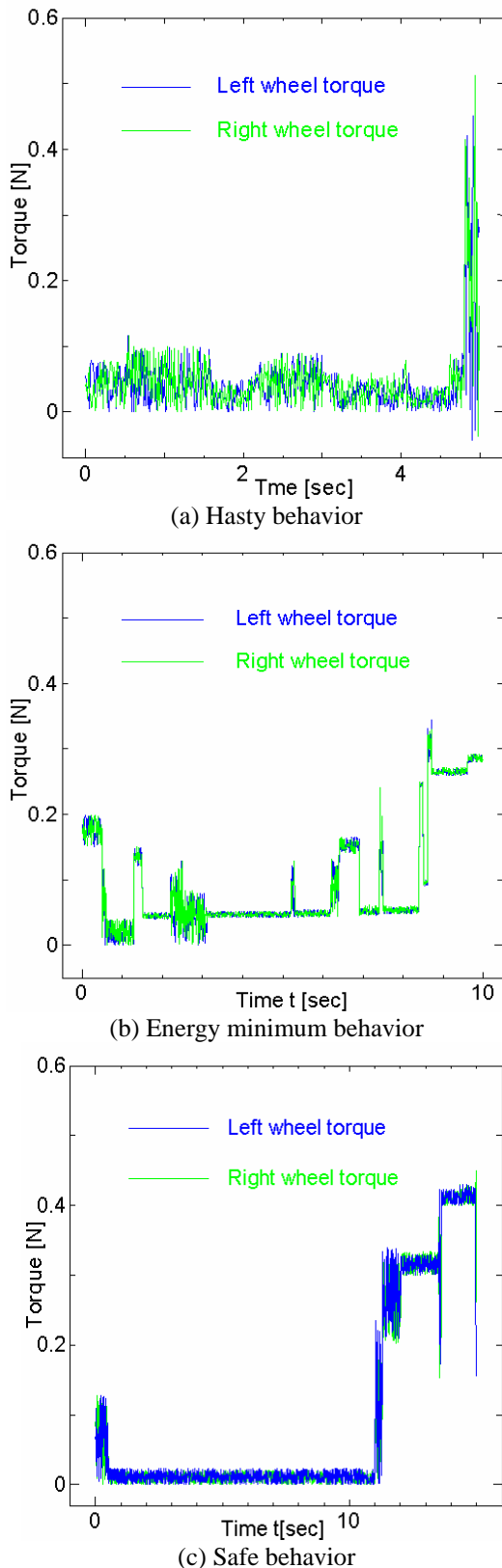
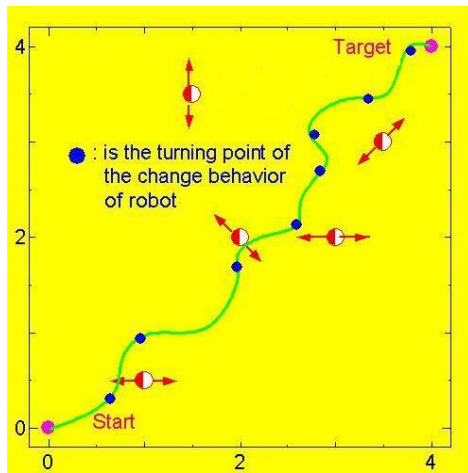


Fig. 12 Torque profiles with fuzzy Q-learning

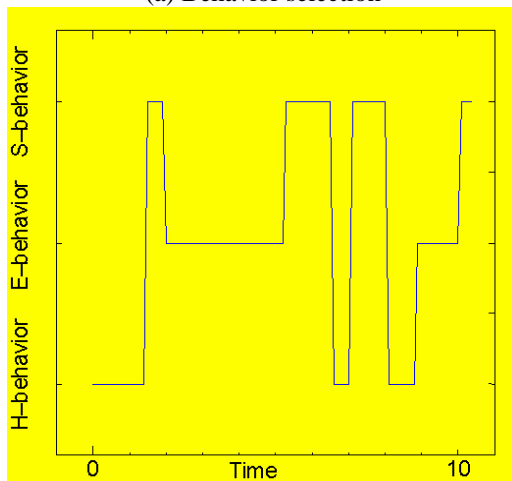
In the simulation result of the hasty behavior with the Q-learning, the robot took 4.53 [sec] from the start point to the target, and the total torque was 172.14 [Nm]. In the simulation result of the energy minimum behav-

ior, the robot took 9.84 [sec] from the start point to the target, and the total torque was 98.74 [Nm]. In the simulation result of the safe behavior, the robot took 18.24 [sec] from the start point to the target, and the total torque was 164.29 [Nm]. In the simulation result of the hasty behavior with the fuzzy Q-learning, the robot took the robot took 4.42 [sec] from the start point to the target, which is about 97.5% of time required in the simulation with Q-learning in the hasty behavior, and the total torque was 160.94 [Nm] which is about 93.5% of torque required in the simulation with Q-learning.. In the simulation result of the energy minimum behavior, the robot took 9.32 [sec] from the start point to the target, which is about 94.7% of time required in the simulation with Q-learning in the energy minimum behavior, and the total torque was 94.86 [Nm] which is about 96% of torque required in the simulation with Q-learning.. In the simulation result of the safe behavior, the robot took 16.42 [sec] from the start point to the target, which is about 90% of time required in the simulation with Q-learning in the safe behavior, and the total torque was 153.45 [Nm] which is about 93.4% of torque required in the simulation with Q-learning. Consequently, one can see that the energy minimum behavior consumes less energy than the other behavior in both results. In hasty behavior, the robot quickly arrives at the target although a lot of energy is consumed. The safe behavior takes a lot of time to get to the target. These results show that the optimal behavior for each situation can be obtained even in the dynamic working environment using the proposed fuzzy Q-learning and ordinary method. However, the proposed fuzzy Q-learning is better than the ordinary method which is Q-learning.

From the simulation results shown in Fig. 13, one can see that the robot effectively selected the certain behavior observing the given states using the proposed behavior selection algorithm in the dynamic working environment.



(a) Behavior selection



(b) Select behavior

Fig. 13 Behavior selection with selection algorithm

6. Conclusions

In this paper, multi-dimensional fuzzy Q-learning has been proposed for the robots to search the optimal behaviors in the dynamic environment with the fuzzy states. The robot performs the optimal behavior by switching the evaluating behavior avoiding the moving obstacles with the fuzzy Q-learning and proposed behavior selection algorithm. In the fuzzy Q-net, the weights and the output are represented by vector form. Each component of vectors is in charge of each evaluation (reward), so that each component of the weight vectors and the output vector of the fuzzy Q-net is adjusted based on fuzzy reward or punishment for each item of the evaluation after a certain behavior is performed. The simulation results showed the effectiveness of the proposed methods.

References

- [1] R.S.Sutton and A.G.Barto, *Reinforcement Learning*, MIT Press, 1998.
- [2] C.J.C.H.Watkins, "Learning from Delayed Rewards," Ph.D. Dissertation, Cambridge University, 1989.
- [3] Y.Arai, T.Fujii, Hajime Asama, H.Kaetsu, I.Endo. "Collision avoidance in multi-robot systems based on multi-layered reinforcement learning," *Robotics and Autonomous Systems* 29, 21-32, 1999.
- [4] S.G. Tzafestas and G.G. Rigatos. "Fuzzy Reinforcement Learning Control for Compliance Tasks of Robotic Manipulators," *IEEE trans. on systems, man, and cybernetics-part: cybernetics*, vol. 32, no.1, February, 2002.
- [5] H.R.Kim, J.H.Hwang and D.S. Kwon. "Human-Robot Cooperation Strategy for Interactive Robot Soccer by Fuzzy Q-learning," *Proc. of the 2003 IEEE/RSJ Intl.Conf. on Intelligent Robots and Systems Las Vegas, Nevada*, October, 2003.
- [6] K.Kiguchi, T.Nanayakkara, K.Watanabe, T.Fukuda, "Multi-Dimensional Reinforcement Learning Using a Vector Q-Net - Application to Mobile Robots," *Internal Journal of Control, Automation and Systems*, vol. 1, no.1, pp.142-148, 2003.
- [7] J.-S.R.Jang C.-T.Sun E.Mizutani, *Neuro-Fuzzy and Soft computing*, Prentice Hall, 1997.
- [8] J.R.Asensio, L.Montano. "A Kinematic and Dynamic Model-Based Motion Controller For Mobile Robots," *Proc. of 2002 IFAC 15th Triennial World Congress, Barcelona, Spain*, 2002.
- [9] D.Gu and H.Hu. "Reinforcement Learning of Fuzzy Logic Controllers for Quadruped Walking Robots," *Proc. of 2002 IFAC 15th Triennial World Congress, Barcelona, Spain*, 2002.
- [10] P.Reignier, V.Hansen, and J. L.Crowley, "Incremental Supervised Learning for Mobile Robot Reactive Control," *Robotics and Autonomous Systems*, 19, pp.247-257, 1997.
- [11] M.Asada, M.Noda, and S.Tawaratumida, "Purposive Behavior Acquisition for a Real Robot by Vision Based Reinforcement learning," *Machine Learning*, 23, pp.279-303, 1996.
- [12] J.Baltes, Y.Lin, "Path Tracking Control of Non-holonomic Car-like Robot with Reinforcement Learning," *Robot Soccer World Cup*, Springer.
- [13] J.del R. Mukkab. "Learning Efficient Reactive Behavioral Sequences form Basic Reflexes in a Goal Directed Autonomous Robot," *Cambridge, MA: MIT Press.*, 1994.
- [14] J. del R. Millan. "Rapid, safe, and incremental learning of navigation strategies," *IEEE Trans. on systems, man, and cybernetics*, vol. 26, no.3, 1996.

[15] Y. Dahmani and A. Benyettou. "Fuzzy Reinforcement Rectilinear Trajectory Learning," *Journal of Applied Sciences*, vol. 4, no.3, pp.388-392, 2004.

[16] C. Ye, N. H.C.Yung, and D. Wang, "A Fuzzy Controller With Supervised Learning Assisted Reinforcement Learning Algorithm for Obstacle Avoidance," *IEEE trans. on systems, man, and cybernetics-part: cybernetics*, vol. 33, no.1, 17-27, February, 2003.

[17] C.T. Lin and I.F. Chung. "A Reinforcement Neuro-Fuzzy Combiner for Multiobjective Control," *IEEE trans. on systems, man, and cybernetics-part: cybernetics*, vol. 29, no.6, December, 1999.

[18] Y.H. Kuo, J.P. Hsu, and C.W. Wang. "A Parallel Fuzzy Inference Model with Distributed Prediction Scheme for Reinforcement Learning," *IEEE trans. on systems, man, and cybernetics-part: cybernetics*, vol. 28, no.2, April, 1998.

[19] L. Jouffe. "Fuzzy Inference System Learning by Reinforcement Methods," *IEEE trans. on systems, man, and cybernetics-part: cybernetics*, vol. 28, no.3, August, 1998.

[20] G. Cicirelli, T. D'Orazio, A. Distante. "Neural Q-learning control architectures for a wall-following behavior," *Proc. of the 2003 IEEE/RSJ intl. Conf. on Intelligent Robots and systems Las Vegas, Nevada*. October, 2003.

[21] A.P.Aguiar, A.N. Atassi, A. M. Pascoal. "Regulation of a Nonholonomic Dynamic Wheeled Mobile Robot with Parametric Modeling Uncertainty using Lyapunov Functions," *Proc. of CDC' 2000-39th IEEE Conference on Decision and Control, Sydney, Australia*, December, 2000.



Kazuo Kiguchi received the Bachelor of Engineering degree in mechanical engineering from Niigata University, Japan in 1986, the Master of Applied Science degree in mechanical engineering from the University of Ottawa, Canada in 1993, and the Doctor of Engineering degree from Nagoya University, Japan in 1997.

He was a Research Engineer with Mazda Motor Co. between 1986-1989, and with MHI Aerospace Systems Co. between 1989-1991. He worked for the Dept. of Industrial and Systems Engineering, Niigata College of Technology, Japan between 1994-1999. He is currently a professor in the Dept. of Advanced Systems Control Engineering, Graduate School of Science and Engineering, Saga University, Japan. He received the J.F.Engelberger Best Paper Award at WAC2000. His research interests include biorobotics, intelligent robots, machine learning, application of soft computing for robot control, and application of robotics in medicine. He is a member of IEEE (R&A, SMC, EMB, IE, and CS Societies), the Japan Society of Mechanical En-

gineers, the Society of Instrument and Control Engineers, the Robotics Society of Japan, the Japan Society of Computer Aided Surgery, the Virtual Reality Society of Japan, and the Japanese Society for Clinical Biomechanics and Related Research.



Hui He received B.E degrees in Automation Engineering from North Eastern University in 2003, and the M.S degree in Mechanical Engineering from the University of Saga, Japan in 2005. He is currently a Ph.D student in the Dept. of Advanced Systems Control Engineering, Saga University, Japan. He is a member of the Japan Society of Mechanical Engineers (JSME) and IEEE.



Kenbu Teramoto received the B.S. and M.S. degrees in engineering and the Ph.D. degree from the University of Tokyo, Japan, in 1983, 1985, and 1988 respectively. Since 1990, he has been an Associate Professor at Saga University, Saga City, Saga Prefecture. Dr. Teramoto is a member of the IEEE Signal Processing Society, the Acoustical Society of America (ASA), the Society of Instrument and Control Engineers(SICE), the Institute of Electronics, Information and Communication Engineers (IEICE), the Acoustic Society of Japan (ASJ), and the Visualization Society of Japan (VSJ).