The 8th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI 2011)
Nov. 23-26, 2011 in Songdo ConventiA, Incheon, Korea

Q Learning Behavior on Autonomous Navigation of Physical Robot

Handy Wicaksono
Department of Electrical Engineering, Petra Christian University, Surabaya, Indonesia
(E-mail: handy@petra.ac.id)

Abstract - Behavior based architecture gives robot fast and reliable action. If there are many behaviors in robot, behavior coordination is needed. Subsumption architecture is behavior coordination method that give quick and robust response. Learning mechanism improve robot's performance in handling uncertainty. Q learning is popular reinforcement learning method that has been used in robot learning because it is simple, convergent and off policy. In this paper, Q learning will be used as learning mechanism for obstacle avoidance behavior in autonomous robot navigation. Learning rate of Q learning affect robot's performance in learning phase. As the result, Q learning algorithm is successfully implemented in a physical robot with its imperfect environment.

Keywords – Q learning, behavior coordination, autonomous navigation, physical robot

1. Introduction

Behavior based architecture is a key concept in creating fast and reliable robot. It replaces deliberative architecture that used by Nilson in Shakey robot [1]. Behavior based robot doesn't need world model to finish its task. The real environment is the only model which needed by robot. Another advantage of this architecture is all behaviors run in parallel, simultaneous, and asynchronous way [2].

In behavior based architecture, robot must have behavior coordinator. First approach suggested by Brooks [2] is Subsumption Architecture that can be classified as competitive method. In this method, there is only one behavior (that can be applied in robot) at one time. It is very simple method and it gives the fast performance result, but it has disavantage of non-smooth response and inaccuracy in robot movement.

In order to anticipate many uncertain things, robot should have learning mechanism. In supervised learning, robot will need a master to teach it. On the other hand, unsupervised learning mechanism will make robot learn by itself. Reinforcement learning is an example of this method, so robot can learn online by accepting reward from its environment [3].

There are many methods to solve reinforcement learning problem. One of most popular methods is Temporal Difference Algorithm, especially Q Learning algorithm [4]. Q Learning advantages are its off-policy characteristic and simple algorithm. It is also convergent in optimal policy. But it can only be used in discrete state/action. If Q table is large enough, algorithm will spend too much time in learning process [5].

Learning algorithm usually takes more memory space on robot's controller and it also adds program complexity than non-learning one. That's why some researchers prefer use this algorithm (including Q learning) on computer simulation only [6 - 8]. However, its implementation on real robot is very important because there are many differences between computer simulation and real world experiment. LEGO NXT robot as low cost and popular robotics kit will used here as a replacement of somewhat expensive research robotics platform.

This paper will describe about Q learning algorithm implementation on physical robot which navigates itself autonomously. Q learning will be applied on single behavior and all behaviors are coordinated by Subsumption Architecture method. This is different approach with Khirji et. al. [9] that used Q learning to coordinate some behaviors

2. Behaviors Coordination Method

In behavior based robotics approach, proper method of behavior coordination is significant. The designer needs to know how robot coordinates its behaviors and take the action in the real world. There are two approaches : competitive and cooperative. In competitive method, at one time, there is only one behavior that applied in robot.

The first suggestion in this method is Subsumption Architecture by Brooks [2]. This method divides behaviors to many levels, where the higher level behavior has higher priorities. So it can subsume the lower level ones. The layered control system figure is given below.

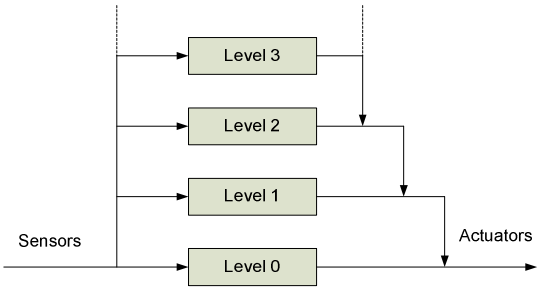


Fig. 1. Layered control system [2]

Robot should have these behaviors to accomplish autonomous navigation task :

1. Wandering

2. Obstacle avoidance
3. Search target
4. Stop

Those behaviors must be coordinated so they can work synchronously in robot. Coordination method which is used in this research is Subsumption Architecture [2]. Figure 2. shows robot's behaviors coordination structure.

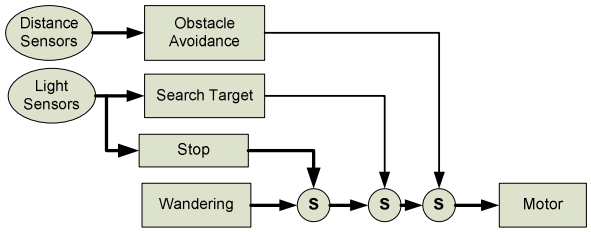


Fig. 2. Subsumption Architecture for autonomous navigation robot

From the figure, it can be seen that Wandering is the lowest level behavior, so if there are another active behaviors, then Wandering won't be active. Behavior with highest priority level is obstacle avoidance (OA).

3. Q Learning

Reinforcement learning is a kind of unsupervised learning method which learns from agent's environment. Agent (such as: robot) will receive reward from its environment. This method is simple and effective for online and fast process in an agent (such as robot). Figure 3. shows reinforcement learning basic scheme.

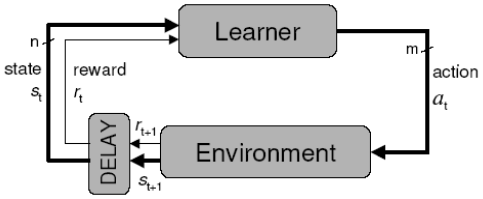


Fig. 3. Reinforcement learning basic scheme [3]

Q learning is most popular reinforcement learning method because it is simple, convergent, and off policy. So it is suitable for real time application such as robot. Q learning algorithm is described in Fig. 4.

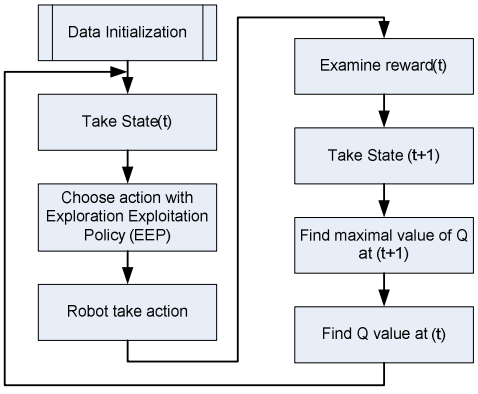


Fig. 4. General flow chart of Q learning algorithm

Simple Q value equation that used in this algoroithm is shown in Eq (1).

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)] \qquad (1)$$

where :

Q(s,a) : component of Q table (state, action)
s : state s' : next state
a : action a' : next action
r : reward α : learning rate γ : discount factor

Design of state and reward are important in Q learning algorithm. Here are states value design of robot's obstacle avoidance behavior :

0 : if obstacle's distance is less then equal with 30 cm from robot's left and right side
1 : if obstacle's distance is less then equal with 30 cm from robot's left side and more than 30 cm from robot's right side
2 : if obstacle's distance is less then equal with 30 cm from robot's right side and more than 30 cm from robot's left side
3 : if obstacle is more than 30 cm from robot's left and right side

Meanwhile rewards value design of the same behavior are :

-2 : if obstacle's distance is less then equal with 20 cm from robot's left and right side
-1 : if obstacle's distance is less then equal with 20 cm from robot's left side or right side
2 : if obstacle obstacle is more than 20 cm from robot's left and right side

In this paper, Q learning will be applied on obstacle avoidance behavior only. Figure 5. Shows Q learning behavior implementation on robot's subsumption architecture.

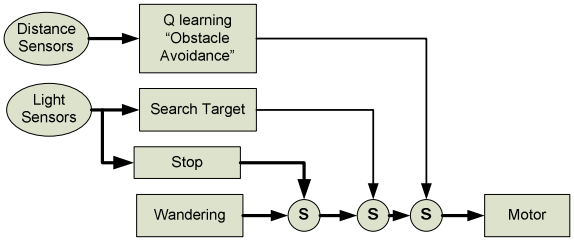


Fig. 5. Q learning behavior on robot's subsumption architecture

4. Physical Robot Implementation

LEGO NXT Robot is a famous robotic kit for people of all ages. It is suitable for developed country like Indonesia because of its affordable price (compare than expensive robotic platform like Kephera, Pioneer, etc). Although its main target user is children and teenager, nowadays LEGO NXT robot has been used in university for advance robotic application such as envioronment mapping [10], multi robot system [11], and robot learning [12].

This paper will describe about implementation of behavior coordination and Q learning on LEGO NXT Robot. NXC (Not eXatcly C), an open source C-like language will be used to program the robot as substitute of NXT-G (original graphical programming tool from LEGO NXT). Its text based programming style is suitable for advance algorithm like Q learning.

There are some NXC programming techniques on implementation of robot's Q learning behavior. Q learning algorithm needs 2 dimensional array to build Q table consist of state action. So enhanced NBC/NXC firmware that support multi dimensional array will be used here. It is also important to use float data type on α (learning rate) and γ (discount rate), so their value can be varied between 0 and 1. Experiment data will be saved on NXT brick as text file and it will be transferred to PC after all experiments are finished.

Robot used in this research has two ultrasonic sensors (to detect the obstacles), two light sensors (to detect the target) and two servo motors. NXT Brick behaves as "brain" or controller for this robot. Figure 6. shows the robot.

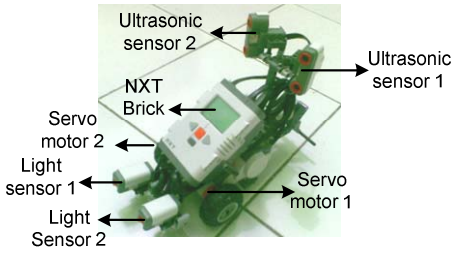


Fig. 6. LEGO NXT Robot for autonomous navigation task

Arena that will be used in experiments have 3 different home positions and 1 target location (by using candle as light source). The general arena is shown in Fig. 7.

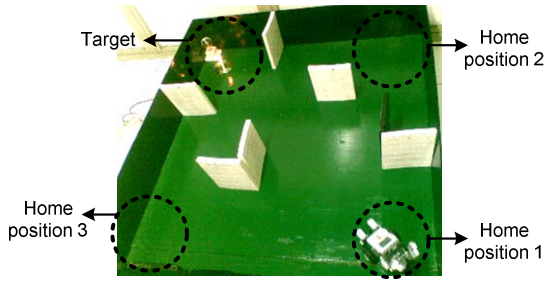


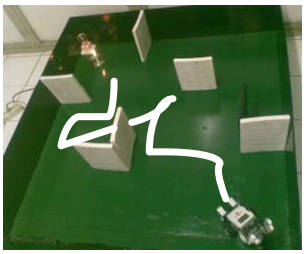
Fig. 7. The arena

Beside this arena, some simple structure of some obstacles and target will also be used in order to know characteristics of learning mechanism clearly.

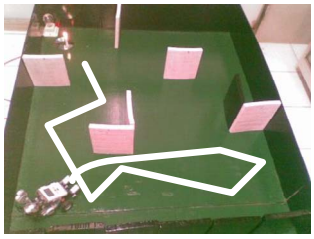
5. Result and Discussion

5.1 Experiment on robot's behaviors coordination

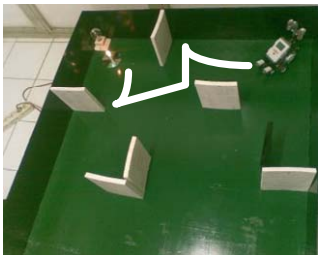
First experiment that will be done is to test robot's ability in solving autonomous navigation task. Given three different home positions, robot should avoid the obstacles and find the target. The result is shown on Fig. 8.



(a)



(b)



(c)

Fig. 8. Robot's trajectory from home position 1, 2 and 3

From Fig. 8. it is obvious that robot with subsumption architecture can avoid the obstacle well. Robot also succeed to find the light source as target from three different home positions.

5.2 Experiment on Q learning behavior with fixed learning rate

As seen on Fig. 5., Q learning only applied in obstacle avoidance behavior. In order to watch robot's performance, a simple obstacle structure is prepared. Q learning algorithm applied on robot use $\alpha = 0.7$ and $\gamma = 0.7$. It utilize greedy method for exploration – exploitation policy. Robot's performance on the beginning and the end of trial is shown on Fig. 9 and Fig. 10.

(a) (b)

Fig. 9. Robot's performance at the beginning and the end of trial 1

(a) (b)

Fig. 10. Robot's performance at the beginning and the end of trial 2

It can be seen from Fig. 9. and Fig. 10. that robot's learning result can be different between one and another experiment. The first robot tend to go to right direction and the second one choose left direction. Both of them are succeed to avoid the obstacle. This can be happened because Q learning give intelligence on each robot to decide which is the best decision (action) for robot itself.

Robot's goal in Q learning point of view is collecting positive rewards as many as possible. Graphic of rewards average every ten iterations and total rewards during the experiment is shown on Fig. 11 and Fig. 12.

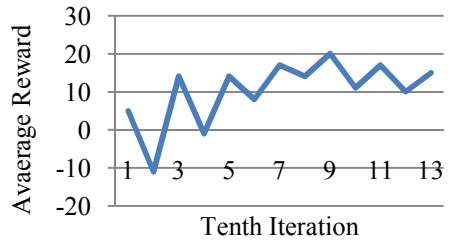


Fig. 11. Average reward every tenth iteration

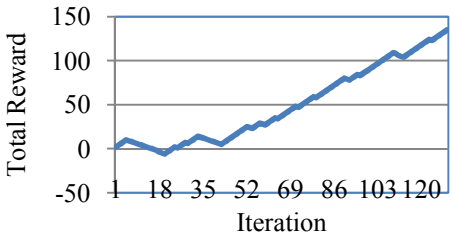


Fig. 12. Total rewards of Q learning obstacle avoidance behavior.

From Fig. 11., it can be seen that average reward that received by robot is getting bigger over the time. In the learning phase robot still receive some negative rewards, but after 5 steps it start to collect positive rewards. Figure 12. Shows total (accumulated) rewards collected by robot is getting larger over the time. So it can be concluded that robot can maximize its reward after learning for some time.

5.3 Experiment on Q learning behavior with varying learning rate

In this experiment, different learning rate (α) will be given to the robot's Q learning algorithm. Its values are : 0.25, 0.5, 0.75 and 1. The result shown in Fig. 13.

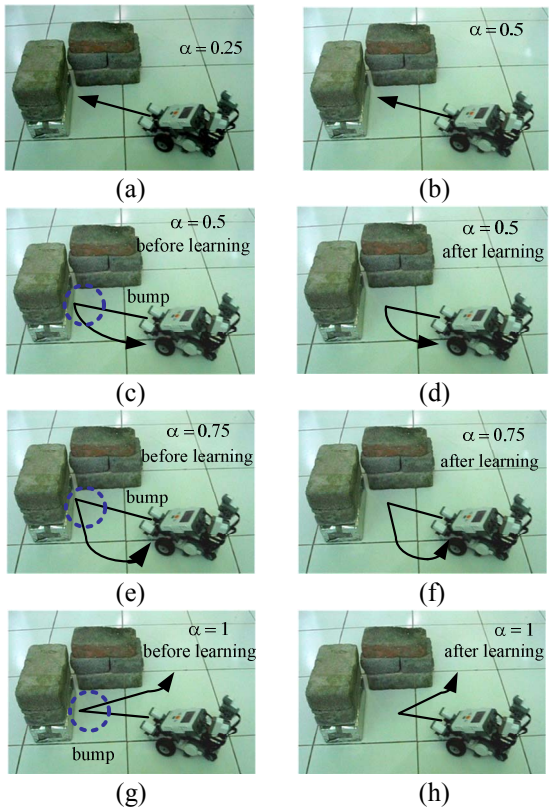


Fig. 13. Robot's movement with different learning rate values

From Fig. 13. (a) and (b), it can be seen that robot with 0.25 learning rate can not learn to avoid obstacles because its value is too small. While robot with 0.5 learning rate sometimes succeed to learn, but it's not happened in every experiment (see Fig. 13. (c) – (d)). But robot with 0.75 and 1 learning rate can learn obstacle avoidance task well everytime (see Fig. 13. (e) – (h)). Before robot learns, it will bump to the obstacles sometime because it still doesn't understand that it is forbidden. But after it has learned, it can avoid obstacle (without bumping) successfully.

The difference of robot with 0.5, 0.75 and 1 learning rate is time needed to learn and finish obstacle avoidance task. Here is the comparison table of them.

Table 1 Comparison of robot with different learning rate.

α	Before learning (seconds)	After learning (seconds)
0.5	15	7
0.75	9	5
1	7	7

From Table 1, it can be seen that the increasing of learning rate is proportional with decreasing time needed by robot to solve the task. In this case, robot with $\alpha = 1$ is the fastest. But in after-learning phase, those robot is not always be the fastest one too.

Beside the time needed to learn and finish the task, rewards that receive by robot with different learning rate is also different. A graph of rewards collected by these robots are shown on Fig. 14.

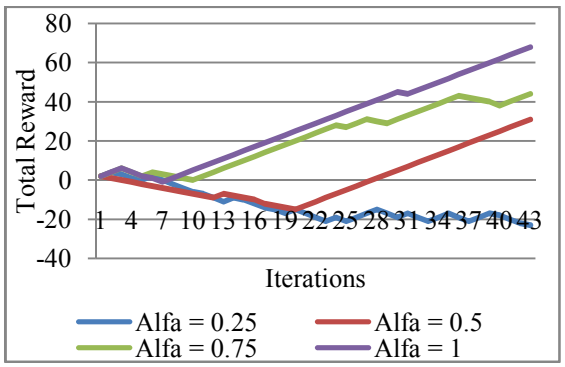


Fig. 14. Total rewards collected by robot's obstacle avoidance behavior.

From figure above, it can be stated that robot with bigger learning rate will collect the bigger amount of rewards too. It means that robot will learn the task faster than the others. So it can be concluded that for simple obstacle avoidance task, the best learning rate (α) that can be given by robot is 1. But it does not always true for every tasks. In some tasks, when a robot learn too fast, it tend to make the robot fall in local minima.

This Q learning behavior has been used in physical robot that solve autonomous navigation task, and it succeed to avoid the obstacle (after some learning time) and reach the target (by its combination with search target behavior). Some problems dealing with imperfect environment should be solved to get the best result.

6. Conclusion

It can be concluded from the experiment results that :

1. Physical robot using subsumption architecture as *behavior coordination* method can finish autonomous navigation task well.
2. Physical robot using Q learning mechanisme can learn and understand obstacle avoidance task well, this is remarked by its success in collecting positive rewards continually.
3. Learning rate of Q learning mechanism affect the robot's learning performance. When learning rate getting bigger, the learning phase getting faster too. But in some tasks, it can drive the robot to fall in local minima phase.
4. Q learning experiments in physical robot give clearer understanding of Q learning algorithm itself, although there is disturbance from the imperfect environment.

Acknowledgement

This work is being supported by DP2M – Directorate General of Higher Education (Indonesia) through "Young Lecturer Research Grant" with contract number 0026/SP2H-PDM/OO7/KL.1/II/2010. Author also thanks Handry Khoswanto for valuable suggestion on LEGO NXT Robot implementation.

References

[1] N. J. Nillson, "Shakey the Robot", Technical Note 323, AI Center, SRI International, 1984

[2] R. Brooks, "A Robust Layered Control System For a Mobile Robot", IEEE Journal of Robotics and Automation, Vol. 2, No. 1, pp. 14 – 23, 1986

[3] R.S. Sutton, and A.G. Barto, Reinforcement Learning, an Introduction, MIT Press, Massachusets, 1998

[4] C. Watkins and P. Dayan, "Q-learning", Technical Note, Machine Learning, Vol 8, 1992, pp.279-292

[5] M.C. Perez, "A Proposal of Behavior Based Control Architecture with Reinforcement Learning for an Autonomous Underwater Robot", Ph.D. Dissertation, University of Girona, Girona, 2003

[6] R. Hafner, and M. Riedmiller, "Reinforcement Learning on a Omnidirectional Mobile Robot", Proceedings of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vol. 1, Las Vegas, 2003, pp. 418 – 423.

[7] H. Wicaksono, Prihastono, K. Anam, S. Kuswadi, R. Effendie, A. Jazidie, I. A. Sulistijono, M. Sampei, "Modified Fuzzy Behavior Coordination for Autonomous Mobile Robot Navigation System", Proc. of ICCAS-SICE, 2009

[8] K. Anam, S. Kuswadi, "Behavior Based Control and Fuzzy Q-Learning For Autonomous Mobile Robot Navigation", Proceeding of The 4th International Conference on Information & Communication Technology and Systems (ICTS), 2008

[9] L. Khriji, F. Touati, K. Benhmed, A.A. Yahmedi, "Q-Learning Based Mobile robot behaviors Coordination", Proc. of International Renewable Energy Congress (IREC), 2010

[10] G. Oliveira , R. Silva , T. Lira , L. P. Reis, "Environment Mapping using the Lego Mindstorms NXT and leJOS NXJ", EPIA, 2009

[11] D. Benedettelli, N. Ceccarelli, A. Garulli, A. Giannitrapani, "Experimental validation of collective circular motion for nonholonomic multi-vehicle systems", Robotics and Autonomous Systems, Vol. 58, No. 8, pp. 1028-1036, 2010

[12] B. R. Leffler, C. R. Mansley, M. L. Littman, "Efficient Learning of Dynamics Models using Terrain Classification", Proceedings of the International Workshop on Evolutionary and Reinforcement Learning for Autonomous Robot Systems, 2008