The Seventeenth International Symposium on Artificial Life and Robotics 2012 (AROB 17th '12),
B-Con Plaza, Beppu, Oita, Japan, January 19-21, 2012

Reduction of learning space by making a choice of sensor information

Yasutaka Kishima[1], Kentarou Kurashige[1], and Toshinobu Numata[1]

[1] Muroran Institute of Technology, Muroran, Hokkaido 050-8585, Japan
(Tel: 81-143-46-5489)
(yasutaka.kishima@gmail.com, kentarou@csse.muroran-it.ac.jp, toshinobu.0122@gmail.com)

Abstract: There are many researches about applying machine learning to robot. Robot uses sensors as input information for learning. When robot performs various task, sensors which are used for each task is important. The reason is important sensors for performing task are different in each task. Robot should use proper sensors for each task. Therefore, we will propose a method which robot can autonomously make a choice of important sensors for each task. We define measure of importance of sensor for task. The measure is coefficient of correlation between sensor value of each sensor and reward on reinforcement learning. Robot decide important sensors based on correlation. Robot reduce learning space based on important sensors. Robot can learn efficiently by reduced learning space.

Keywords: Reduction of learning space, Robotics, Data mining, Reinforcement learning

1 INTRODUCTION

In recent years, robots which have various sensors and actuators are appeared with development of hardware technology. These robots is hoped to use for various task. In order for robot to perform various task, there are many researches about applying machine learning to robot[1]-[3]. Learning for robot is to obtain correspondence between input and output for achieving purpose. The input is information from installed sensors and output is action of robot.

Robot uses sensors as input for learning. The way of use of sensors is important for task. When robot is used for various tasks, installed sensors which are used becomes a problem. The problem is that learning efficiency can be down depend on the way of use of installed sensors in some tasks. Because important sensors for performing task are different. Robot has important sensors and unimportant sensors for a given task. Input of robot is increase for unimportant sensors. When input is increase, the number of corresponding action which agent should learn to input increases. Therefore, learning time increases by using unimportant sensors for learning. As the result learning efficiency is down.

Therefore, robot should learn by using important sensors only for efficient learning. To achieve this, robot should have ability which robot makes a choice of important sensors for task. Our purpose is to propose a system which robot can make a choice of important sensors for any task and robot can learn efficiently by using information from important sensors only.

This system is effective to various task. Robot can make a choice of important sensor for changed task by this system. And information which robot should learn, which is correspondence between information from sensors and action is reduced. As the result, robot can learn efficiently.

2 CONCEPT OF MAKING A CHOICE OF IMPORTANT SENSORS

To make a choice of important sensor, robot needs measure of importance of each sensor for task. We focus on correlation sensor value and degree of achievement of task as the measure of importance of each sensor for task. Many tasks have correlation sensor value and degree of achievement of task. For example, in garbage collection task robot need to approach a garbage to pick up it. Degree of achievement of task is shown distance between the robot and the garbage, and increases as the robot approaches the garbage. There is a correlation between the degree of achievement of the task and distance between robot and the garbage like this.

We show outline on fig.1. Fig.1 shows the case which robot has 2 kinds of sensors, but robot can have more than 2 kinds of sensors. This robot has sensors and ability to act. The robot recognizes environment around the robot by installed sensors. Environment around the robot expressed group of sensor value of each sensor. The robot collects sensor value of each sensor and degree of achievement of task.

When robot act, sensor value of each sensor is changes. Therefore, robot can collect sensor value of each sensor by repeating action. Robot also collects degree of achievement of task. Robot estimates degree of achievement of task based on change of sensor value of each sensor. Environmental information which robot recognizes changes before and after action. Therefore, robot can know whether changed sensor value is good of not for performing task. In example of garbage collection task, focusing on distance sensor for garbage, when move for a garbage, the distance sensor value reduces. When the robot recognize the reduced distance sensor value, the robot can know current distance sensor value is good for performing the garbage collection task.

The robot finds out correlation between sensor value of each sensor and degrees of achievement of task. In fig.1, robot finds out correlation about sensor 1 and sensor 2. There are negative correlation and positive correlation. The robot estimates important sensors which has negative or positive correlation.

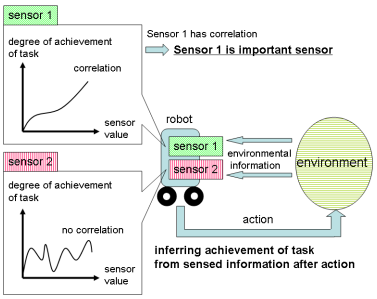

Fig. 1. The outline of making a choice of important sensors

3 PROPOSED SYSTEM ON REINFORCEMENT LEARNING

3.1 Outline of proposed system

The proposed method are used for efficient learning. Therefore, the proposed method is used in combination with a machine learning method. In this paper, we apply reinforcement learning as learning method(RL)[4]. The reason of applying RL is often adopted for experimental robots[5].

We show the system we propose in fig.2. In fig.2, the proposed system is divided into two part. One is the proposed method which robot makes a choice of important sensors. Another is reinforcement learning.

In the proposed method part, robot makes a choice of important sensor for task based on correlations between each sensor and degree of achievement of task. In RL, degree of achievement of task is shown as reward. Robot collects sensor value of each sensor and reward. Robot calculates correlation between each sensor and reward. Then robot decide important sensors based on correlation.

In reinforcement learning, robot learns proper action for task. In this part, there are action evaluation part, creating temporary Q-space based on important sensors part, and action selection part. Action evaluation part is updating evaluation for pair of state and action. State is composed of sensor value of each sensor. In creating temporary Q-space based on important sensors part, robot creates temporary Q-space which is composed of important sensor only. In action selection part, robot select action for recognized state by sensor based on temporary Q-space.

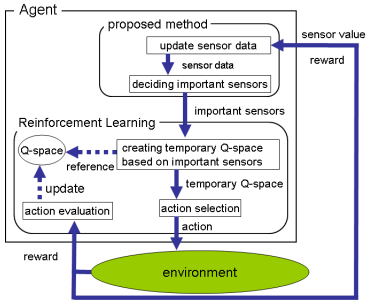

Fig. 2. The outline of proposed system

3.2 Decision of important sensors based on correlation

In proposed method, robot decides important sensors based on correlation between sensor value of each sensor and reward. Robot has sensor value of each sensor and reward which robot has experienced as knowledge. But reward is given for state-action pair. Therefore, there are some reward pattern for a state by an increment of the number of action. In this study, rewards for the state where robot experienced are averaged. Averaged reward at state s is r'_s.

Robot stores sensor value of each sensor and averaged reward as two lists in fig.3 and fig.4. There are state id and sensor value of each sensor in experienced states list. The state id is identification number. The list in fig.3 is an example of experienced states list. There are state id and averaged reward in averaged reward list. The list in fig.4 is an example of averaged reward list.

When robot recognizes state which is not in experienced states list, the robot adds the recognized state. Then robot calculates r'_s and added r'_s, to the averaged reward list. When recognized state is in the experienced states list, robot calculates r'_s and updates r'_s at recognized state in the averaged reward list.

state id	sensor value of sensor 1	sensor value of sensor 2	• • •	sensor value of sensor n
1	s	56	• • •	111
2	23	45	• • •	123
⋮				
m	90	10	• • •	15

Fig. 3. The experienced states list

state id	r'_s
1	111
2	123
⋮	
m	15

Fig. 4. The averaged reward list

Robot calculates coefficient of correlation based on the experienced states list and the averaged reward list. C_j as coefficient of correlation of sensor j and r'_i is calculated by eq.(1).

$$C_j = \frac{\sum_{j=1}^{m}(s_{i,j} - \bar{s}_i)(r'_{,i} - \bar{r'}_i)}{\sqrt{\sum_{j=1}^{m}(s_{i,j} - \bar{s}_i)^2}\sqrt{\sum_{j=1}^{m}(r_{j-\bar{r'}})^2}} \quad (1)$$

The i is identifying number of sensor. The $s_{i,j}$ is sensor value

at state id i and sensor id j at the experienced states list. The r_j is averaged reward at state id j in the averaged reward list. The C_i is calculated for each sensor every action. Robot decides important sensor by comparing $|C_i|$ and Th. The Th is threshold to judge important sensor. When the $|C_i|$ is over Th, robot judges sensor as important sensor.

3.3 The way of use of important sensor for learning

Important sensors are used for selection of action in reinforcement learning. Creating temporary Q-space which is composed only important sensor axes by reduction for unimportant sensor axes. Reduction for Q-space is learning space. Reduction for Q-space is to project Q-values in unimportant sensor axes to Q-values in important sensor axes. Fig.5 shows an example of degeneration of Q-space. Important sensor is S_1, unimportant sensor is S_2 in this example. Q-values in S_2 is projected to Q-values in S_1. Projection is addition Q-values in S_2 to Q-values in S_1.

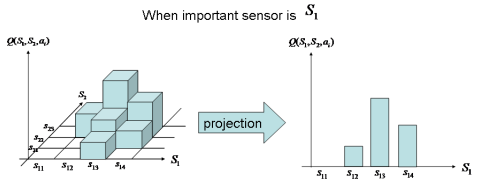

Fig. 5. The projection for temporary Q-space

We formulate based on this outline. Q-value is averaged reward by the number of experience of state-action pair. The number of experience for each state-action pair can be difference. This difference means difference of confidence of Q-value for state-action pair. Therefore, we consider weighted averaging Q-value according to the number of experience. To do this, we estimate total reward based on Q-value and the number of experience for state-action pair. Temporary Q-space is created based on total reward and the number of experience for state-action pair.

Temporary Q-space is used for selection of action at current state. Therefore, we consider projection at only current state of robot. Current state of robot is shown as $S = e_{1,f}, e_{2,g} \ldots, e_{n,z}$. The $e_{1,f}$ means f th sensor value in sensor 1. States of important sensors are shown as $S^\star = e_{1,f}, e_{2,g} \ldots . e_{p,k}$. States of unimportant sensors are shown as $U = e_{p+1,x}, e_{p+2,y} \ldots e_{n,z}$. Total reward of any action at state S ($R(S,a)$) is defined as eq.(2).

$$R(S,a) = R(S^\star, U, a) = Q(S,a) \times E(S,a) \qquad (2)$$

The temporary Q-value $Q(S^\star, a)$ is defined as eq.(3). The $E(C,U,a)$ is the number of experience of state action pair $(S^\star, U, a)$.

$$Q(S^\star, a) = \frac{\sum_x \sum_y \ldots \sum_z R_{total}(C,U,a)}{\sum_x \sum_y \ldots \sum_z E(C,U,a)} \qquad (3)$$

We apply ϵ-greedy method for selecting action. The ϵ-greedy method selects the action which has the highest Q-value basically in current state S. But the method selects an action randomly with probability ϵ.

3.4 The action evaluation

We apply weighted averaging method as action evaluation method. The weighted averaging method evaluates by giving weight to reward which robot obtains recently. When current state of robot is S and the selected action is a, Q-value($Q(S,a)$) is updated by eq.(4). The $\alpha$is step size parameter ($0 \leq \alpha \leq 1$).

$$Q(S,a) \leftarrow Q(S,a) + \alpha\,[r - Q(S,a)] \qquad (4)$$

4 EXPERIMENT TO CONFIRM EFFECTIVENESS OF THE PROPOSED SYSTEM

4.1 Outline of experiment

In this section, we examine effectiveness of the proposed system on computer simulation. Environment for the experiment is shown fig.6. This environment is grid field u × u which is surrounded by walls. Length of both walls is u squares. We prepare virtual robot which is called agent. Agent has two distance sensors. One can measure distance between current position of agent and wall A in fig.6. Another can measure distance between current position of agent and wall B. Agent can move up, down, right and left.

In this experiment, agent performs two kinds of task. One is to depart from wall A. Another is to depart from wall A and wall B. Agent can obtain reward according to distance from wall A or both walls. Agent is allocated upper left of environment. When agent move n_{act} times, task is finished.

We confirm effectiveness of the proposed system by comparing the proposed system and ordinary RL. Therefore we prepare two types of agents to compare. One is applied the proposed method. Another is applied the RL only. In the case of the RL only, robot uses all installed sensors. We compare these agents in obtained total reward after moving n_{act} times.

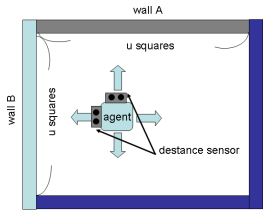

Fig. 6. The environment for the experiment

4.2 The setting of this experiment

We explain about reward for each task and show parameter settings. First, we explain about reward for each task.

In task A, agent can obtain higher reward with increasing the distance between current position of agent and wall A(d_A). Reward of task A is defined in eq.(5).

$$r = d_A \tag{5}$$

In task B, agent can obtain reward based on d_A and the distance between current position of agent and wall B (d_B). Agent can obtain higher reward with increasing both d_A and d_B. Reward of task B is defined in eq.(6)

$$r = d_A + d_B \tag{6}$$

We show parameter settings of this experiment at table.1.

Table 1. The settings of experiment

u	20
n_{act}	20000
ϵ	0.1
α	0.7
initial value of Q-value	0
Th	0.6

4.3 Result and consideration of the experiment

We show result of experiment fig.7 and fig.8. First, we discuss the result about task A. Fig.7 shows obtained reward of agent at each action. The proposed method shows higher convergent of reward than reinforcement learning under 5000 actions. The reason is that robot could select proper sensor for learning by the proposed method. Temporary Q-space has more information than Q-space because of projection of Q-space. The agent using temporary Q-space was easy to select proper action for recognized state.

Next, we discuss the result about task B. Fig.8 shows obtained reward of agent at each action. The proposed method was unstable in early phase of learning(under 5000 actions). The reason is that agent which was applied the proposed method spent times to estimate important sensor. Reward in task B is depend on d_A and d_B. Agent need to collect more sensor and reward data than the case of task A. In early phase of learning, agent selected unimportant sensor because of a lack of sensor and reward data. Therefore reward of agent which was applied the proposed method was lower than agent which was applied reinforcement learning only.

From these result, the proposed method is effective in the case of task which agent uses a part of installed sensor of agent. In the case of task which agent uses all installed sensors of agent, performance of learning in the case of the proposed method is lower than in the case of use of all installed sensors. But agent which is applied the proposed system can change sensors according to importance of each sensor to a task. The agent can learn faster depending on task.

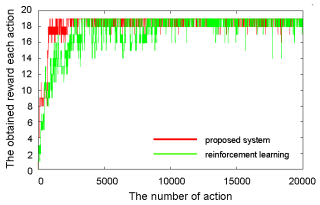

Fig. 7. Obtained reward at each action in task A

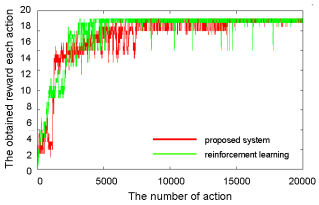

Fig. 8. Obtained reward at each action in task A

5 CONCLUSION

In this paper, we proposed the method which robot makes a choice of important sensors for a task. We constructed a system which is composed proposed method and reinforcement learning. This system is that robot can learn efficiently by the proposed method. We examined effectiveness of the constructed system on simulation. From result of simulation, we confirmed effectiveness of the constructed system and the proposed method. We confirmed robot can learn efficiently by making a choice of important sensors. We will attempt to install this proposed system into actual robot and examine effectiveness as our future work.

REFERENCES

[1] K.merrick, Motivated Learning from Interesting Events: Adaptive, Multitask Learning Agents for Complex Environments, Adaptive Behavior vol.17 no.1 2009, pp.7-27.

[2] Pardowitz.M, Knoop.S, Dillmann.R, .Zollner.R.D,Incremental Learning of Tasks From User Demonstrations, Past Experiences, and Vocal Comments, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on vol.7 No.2 pp.322-332, 2007.

[3] Y.Miyazaki, K.Kurashige, Use of reward - independent knowledge on reinforcement learning for dynamic environment, Proc. of the International Conference on Advanced Computer Science and Information Systems, pp.303-309, Bali, Indonesia, 20th - 23rd Nov. 2010

[4] R.Sutton, A.Barto, Reinforcement Learning, The MIT Press, 1998.

[5] Kober J Person, Oztop E and Peters J, Reinforcement Learning to Adjust Robot Movements to New Situations, Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, pp.2650-2655, 2010.