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

The Application of RL and SVMs to Decide Action of Mobile Robot

Kwang-won Ko, Yong-sul Oh, Qeun-yong Jung and Hoon Heo
Department of Control and Instrumentation Engineering, Korea University
1, 5-ka, Anam-dong, Sungbuk-ku Seoul 136-701, Korea
Email: komb@korea.ac.kr

Abstract - **Support Vector Machines (SVMs) is applied to a practical problem as one of standard tools for machine learning. The application of Reinforcement Learning (RL) and SVMs in action of mobile robot is investigated. A technique to decide the action of autonomous mobile robot in practice is explained in the paper. The proposed method is to find a basis for good action of the system under unknown environment. In multi-dimensional sensor input, the most reasonable action can be automatically decided in each state by RL. Using SVMs, not only optimal decision policy but also generalized state in unknown environment is obtained.**

Key Word: SVM, RL

I. Introduction

We often regard that environment is well known. Actually it is not. Nevertheless, life must go on under unknown circumstances. The living things have been well adapting to the nature. But there has been no teacher for the way to adapt to the nature. They have learned by interacting with environment. Reinforcement Learning is learning how to decide action. It is different from supervised learning. It needs not a teacher. And also it has some characteristics. They are trial-and-error search, delayed reward and so on. Representing method of reinforcement learning is Q-Learning. In the paper, mobile robot learns by using Q-Learning. The Result of Q-Learning is discrete.

There are various methods to make discrete state be continuous. For example, there are Gradient-Descent Method and Tile-Coding. In the paper, discrete state that is solved by reinforcement learning is generalized by using SVM. And this approach has been successfully applied. [1]

In section 2, Q-Learning is described as one of the Reinforcement Learning is explained. In section 3, SVM method and how to set the environment and learning method are described. In section 4, experiment of mobile robot is illustrated. In section 5 results are shown. Conclusion and discussion are presented in section 6.

II. Reinforcement Learning

A. Reinforcement Learning

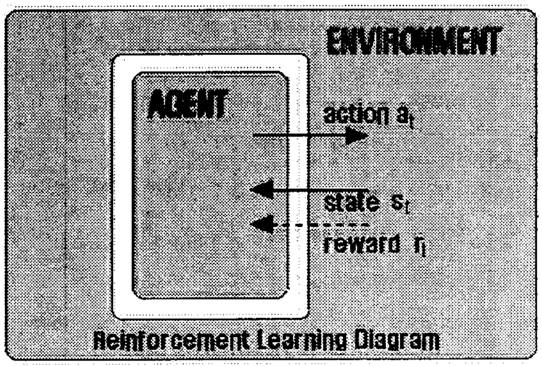

Fig. 1 Reinforcement Learning

Reinforcement Learning is learning from interaction between agent and environment as in Fig. 1. The learner and decision maker is regarded as 'the agent'. 'The environment' is everything outside the agent. The agent take actions and the environment responding to those actions, then it will change new situations to the agent. The environment also gives rewards and new state. If the agent maximizes rewards, then its task can be defined as successfully completed to our goal.

B. Q- Learning

Q-Learning is reinforcement learning algorithm to get value of State-Action, which is to act in the best fitted.

State-Action Value is defined as follows.

$$Q^{\pi}(s,a) = E_{\pi}\{R_t \mid s_t = s\,,\, a_t = a\} \tag{1}$$

$$\text{where } R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$$

Q-Learning Algorithm is as follows [3].

```
Initialize Q(s, a) arbitrarily
Repeat(for each episode):
  Initialize s, a
  Repeat (for each step of episode)
  Choose a' from s'
  Take action a, observe r, s'
 Q(s_t,a_t) <- Q(s_t,a_t) + α [r_{t+1} + γ max_a Q(s_{t+1},a) - Q(s_t,a_t)]
  s <- s';
 until s is terminal
```

Table 1. RL Algorithm

By the repetition of the above Algorithm, the optimized solution can be obtained.

III. SVM

SVM (Support Vector Machines) is kind of the neural networks[4]. SVM can classify data by the subset $d_i = +1$ and the subset $d_i = -1$.

Consider the training sample:

$$\{(x_i\,, d_i)\}_{i=1}^{N} \tag{2}$$

where x_i is the input data, d_i is the target output.

The equation of a decision surface in the from of a hyper plane that does the separation is :

$$w^T x + b = 0 \tag{3}$$

In classification, the pair (w_o, d_0) must satisfy the constraint

$$\begin{aligned} w_o^T x_i + b_o &\geq 1 \qquad && for\ d_i = +1 \\ w_o^T x_i + b_o &\leq -1 \qquad && for\ d_i = -1 \end{aligned} \tag{4}$$

$$where\ w_o, b_o\ is\ optimal\ values$$

Then, we find that they satisfy the constraints

$$d_i(w^T x_i + b) \geq 1 \qquad for\ i = 1,2,\ldots,N \tag{5}$$

We have to minimize w. And outliers trade off these importance.

Therefore we obtain optimize problem as below:

Trade-off Constant $C > 0$

$$\min\ \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i \tag{6}$$

$$subject\ to\quad d_i(< w, x_i > +b) \geq 1 - \xi_i$$

$$\xi_i \geq 0\,,\quad i = 1,\quad ,N$$

The dual problem given by

$$\max\ \theta(\alpha) = \sum \alpha_i - \frac{1}{2} \sum_{i=0}^{N} \sum_{j=0}^{N} \alpha_i \alpha_j d_i d_j k(x_i, x_j) \tag{7}$$

$$subject\ to\quad \sum_{i=1}^{N} \alpha_i d_i = 0, \qquad 0 \leq \alpha_i \leq C,\ \ \forall i.$$

IV. EXPERRIMENT

A. Mobile Robot

The Mobile Robot consists of three parts. The first part is sensor controller using controller, 90S8515 with 8-channel ADC (Analog-Digital Converter). There are five sensors: Which are

- Three IR Sensors
- One Temperature Sensor
- One Front Touch Sensor

The second and the third parts are motor controller using 90S2313. The motor system consists of motor, headgear and encoder. The motor controller can control distance from the encoder signal.

All of three parts are connected to the main computer by serial communication port (COM1). They can transfer sensing data and also receive command from main computer.

B. Environment

The mobile robot does not have any prior knowledge about the environment. The mobile robot only recognizes after it contacts or sense any change in temperature.

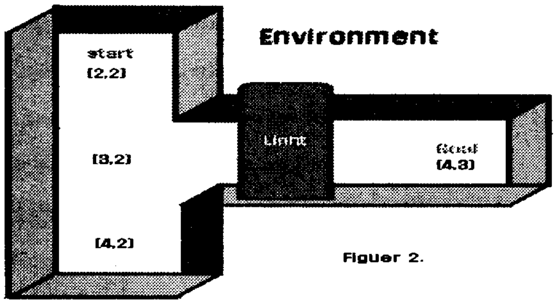

Fig. 2 Experimental Environment

The environment consists of wall and light, as is Fig.2, 3. The mobile robot is constrained by wall. It cannot cross wall. But the mobile robot can cross light with some bad reward.

C. Concept

At first, the main computer initializes states. And then the goal is defined. No prior knowledge about the number of state in the environment is given to the system. So, the main computer assigns current location. When the mobile robot explore environment, it can find new state. Then the main computer updates environment map.

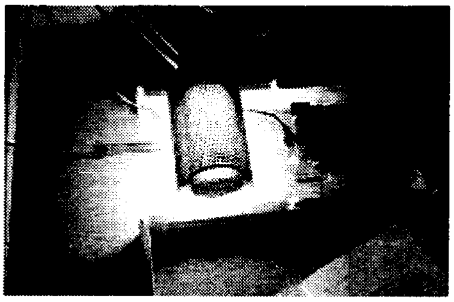

Fig. 3 Experiment

The mobile robot collects data about environment, which is processed in the main computer. After one step, the robot gathers data, and transfers it to computer. It update Q value each step. After one episode that is achieved the goal, Q values are generalized by using SVM.

V. RESULT

A. Q-Learning

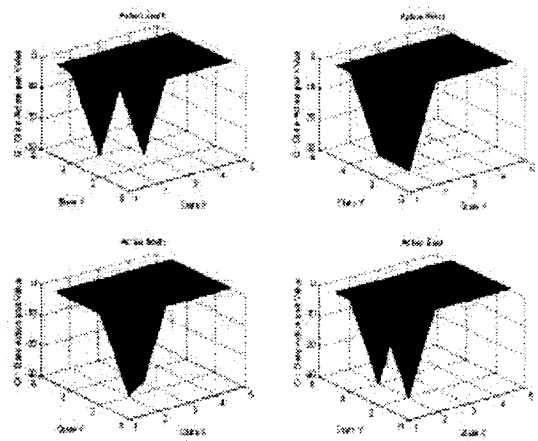

Fig. 4 Optimized State-Action Value

Fig.4 shows optimized value of State-Action. It reveals that each State-Action value has been well learned. It also can be seen that the value of action to the blocked area in the environment is significantly low.

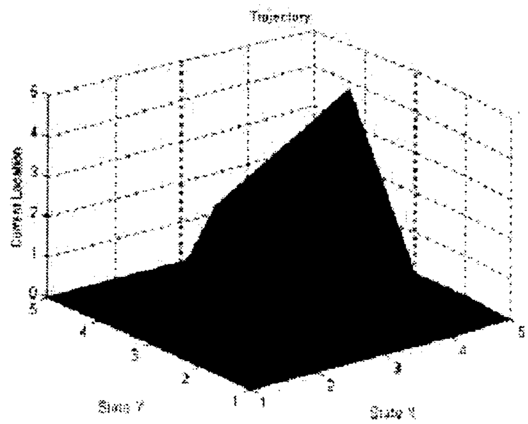

Fig. 5 Optimized Trajectory

Fig.5 shows optimized trajectory. As can be seen, the optimal policy has been found.

B. Generalization

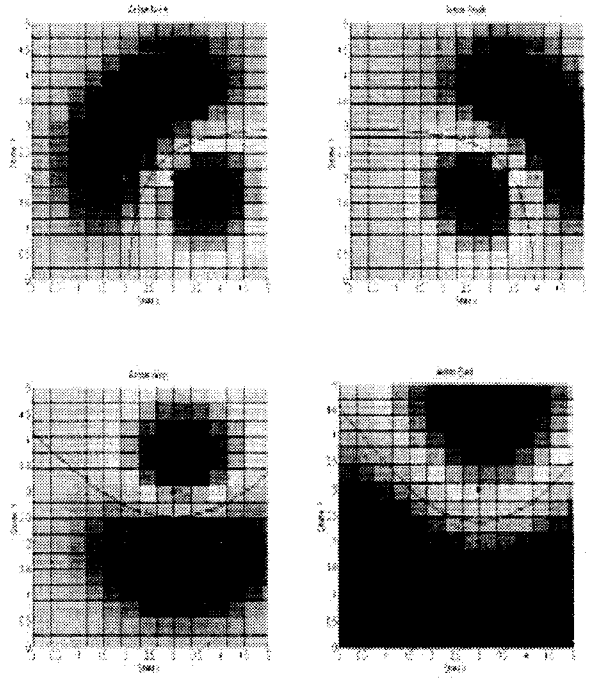

Fig. 6 Generalized State-Action

Figure 6 shows Generalized State-Action value. Degree of action to be taken can be seen in every spot. Accordingly the information on the action to take is obtained when we have SVM solution.

Figure 7 shows classified sensor data. The collected information can be used as a standard of deciding 'good' or 'bad'.

VI. CONCLUSION AND DISCUSSION

In the paper simple application of RL and SVM in combined manner is introduced. When no prior-knowledge about the environment is available, RL is a useful tool. However it requires too many exploring procedure. In the study, the Q-value is approximated without visiting every state by SVM generalization, which shows practically good and promising performances.

Quantitative refining work on the proposed technique is under way.

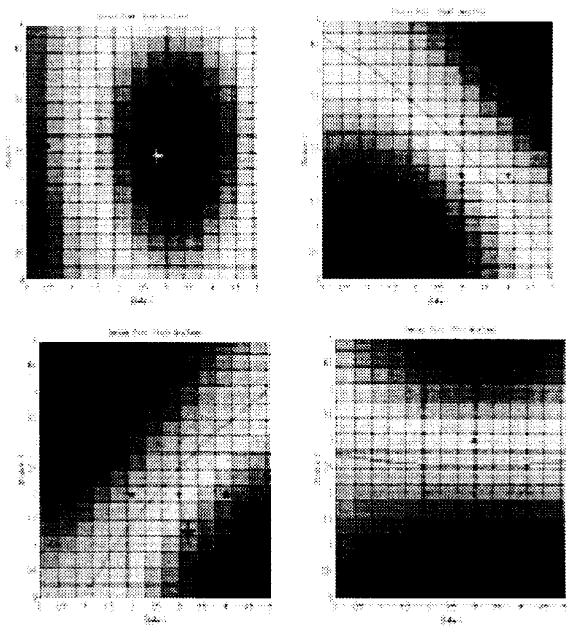

Fig. 7 Classifying the Sensor Data

It can be concluded that this idea can be implemented to autonomous vehicle, which is supposed to work in unknown, often hostile environment such as space exploring vehicle whose mission is on least known planet.

REFERENCE

[1] Ryo Goto, Toshihiro Matsui and Hiroshi Matsuo "State Generalization with Support Vector Machines in Reinforcement Learning", 4th Asia-Pacific Conference on Simulated Evaluation and Learning, Vol.I, pp.51-55 (2002/11)

[2] M.Asada, S.Noda, and K. Hosoda, "Action-Based Sensor Space Categorization for Robot Learning" , in Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems pp.1502-1509, 1996

[3] Richard S. Sutton and Andrew G. Bato "Reinforcement Learning -An Introduction" The MIT Press 1999

[4] Simon Haykin "Neural Networks -a comprehensive foundation" Prentice Hall, 1999

[5] Christopher M. Bishop "Neural Networks for Pattern Recognition" Oxford University Press, 1995