The Fifteenth International Symposium on Artificial Life and Robotics 2010 (AROB 15th '10),
B-Con Plaza, Beppu,Oita, Japan, February 4-6, 2010

Reinforcement Learning Using Voronoi Space Division

KATHY THI AUNG Takayasu FUCHIDA

kathythiaung@gmail.com fuchida@ibe.kagoshima-u.ac.jp

Information and Computer Science Department, Engineering Faculty, Kagoshima University, Kagoshima, Japan
(Tel: 81-99-285-8453; Fax: 81-99-285-8464)

Abstract: This research is concerned with the study method of the robot and program. It deals with the reinfo rcement learning and has gradually become one of the most active research areas in machine learning, artificia l intelligence and neural network research [3]. Using the Voronoi diagram for space division of reinforcement learning creates a new Voronoi region which permits an arbitrary point in the plane. This paper constructs a new
method for space division by a Voronoi diagram and shows some results in four-dimensional spaces.

Keywords: Reinforcement Learning, Q-Learning, Q Element, Voronoi Diagram, VQE

I. INTRODUCTION

Q-learning is preserved in the table, it has the value of each action in a certain state as action value (Q value) and it is called Q table [4] but the size of Q table will increase rapidly by Curse of dimensionality when continuous is treated by Q table and it is not realistic.

Curse of dimensionality means if the dimension number of state space increases, the state also increases exponentially when state space is divided like lattice and the learning speed decreases dramatically. [4]

II. PREVIOUS STUDIES IN THE FIELD

We introduced table element (QE) that dynamically gives the Q value to the state action pair. It is difficult to update of adaptive QE hand it to the state space because of it makes the shape of QE into a rectangular QE. The problem area of adaptive QE is difficult to add, divide and integrate. To solve this problem, we propose VQE using the concept of Voronoi division.

III. PURPOSE

The focus of this study is Voronoi diagrams for space partition creates Voronoi regions where Voronoi Q-value Elements (VQE) are to be located in the continuous state space and it has been proposed for solving the waste of spaces.

The present paper proposes applying a Voronoi division for the continuous state space satisfying the above conditions. Concerning with these region, a Voronoi diagram are defined.

IV. RESEARCH METHODS

1. Reinforcement Learning

Reinforcement learning (RL) is one of the most important learning methods for machine learning and is learning by interacting with an environment. [1] In a reinforcement learning paradigm, a system called agent repeatedly observes the current state of its environment, and then select an action and performs an action. The environment changes the state by the agent's behavior and returns the reward. The agent transits the next state and takes the reward. And then learns based on the reward and updates by its own action policy. [1] Figure 6 shows the fundamental standard feature of framework for RL.

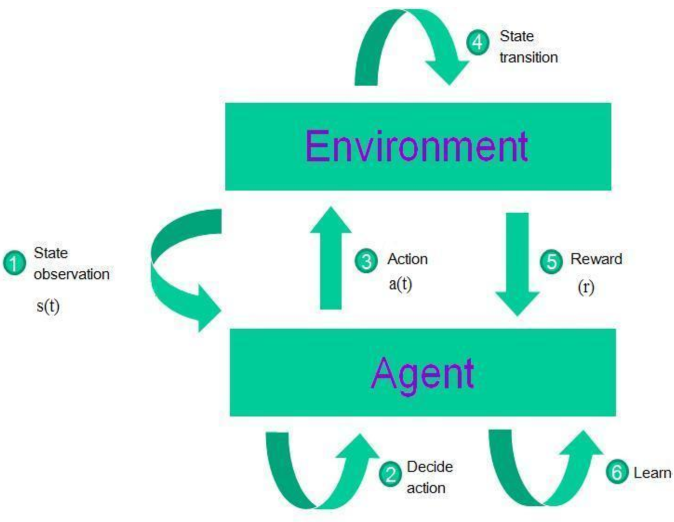

Fig.1. Reinforcement Learning

A. Q-learning

Q-learning is one of the typical techniques of reinforcement learning (RL) and has been applied in this study [1]. The Q-function is used to predict the discounted cumulative reinforcement (also called Q value) for each state-action pair (*s*, *a*) given that the agent is in that state and executes that action because of the agent learns a mapping from states and actions to their Q values. We can estimate the Q values using this method because an action can be chosen just by taking the one with the maximum Q value for the current state.

At each step state (*s*), choose the action (*a*) which maximizes the function $Q(s_t, a_t)$ where Q is the estimated utility function that it tells us how good an action is given a certain state and execute it. And then it receive immediate reward (*r*) and observe the new state (s_{t+1}). $Q(s_t, a_t)$ is immediate reward for making an action and best utility (*Q*) for the resulting state.

The Q-function can update in the following way after looking farther ahead:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha(r + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t))$$

This equation means if the Q value of next state is higher than the current Q value, it depreciate the current Q value. On the other hand, if low, upgrade the current Q value.

Where γ is a discount factor and α is the learning rate. $Q(s_t, a_t)$ is time(*t*), state(*s*), action(*a*) of Q value.

B. Voronoi Diagram (VD)

Voronoi diagram (VD) has the property that for each site (clicked with the mouse) every point in the region around that site is closer to that site than to any of the other sites. [2] Generally, Voronoi like partitions. It can be used to preserve spatial relationships. [2] But if space is metric space, where distance between a pair of points can be defined but if space is as set of dimensions that are not related spatial. In dimension, constructing the Voronoi diagram (VD) is no longer the method of choice for finding the closest pair, due to its exponentially increasing size but it introduced for finding the nearest neighbors of a point in expected time. So it can solve the all nearest neighbor problem for each point and can find the shortest paths. The graphic of VD is illustrated in figure 2, where the color areas indicate the Voronoi diagram belong the black point and the black point are the mother point arranged in space.

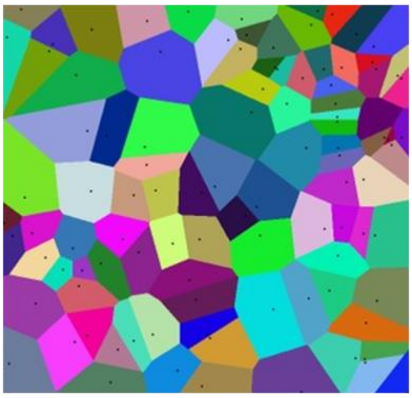

Fig.2. Voronoi Diagram

2. Voronoi Space Division of State Space

A. Division methods of Q-learning

The normal Q-learning is suit for the discrete space but it is difficult to treat the continuous state space. So when we treat such continuous cases, it needs to divide the space into a lot of smaller discrete regions. Using the normal Q-learning in continuous state space, the state space is divided like lattice but the lattice like division possibly generates the waste spaces. So when we divide the normal Q learning in continuous state space by lattice method, it is evenly arranged the Q value in all over the space and if the dimension increases, the number of state also increases by exponentially.

B. VQE (Voronoi Q-value Elements)

Learning evaluates the necessary state in details but the unnecessary part of state make up as a method of roughly so it suggested to use the area of division technique that is called Voronoi diagram (fig2) and it partitioned which arbitrary point is nearest compared with mother points arranged on a space.

In the case of the input point enters within more than one of the region, the input point and each of VQE of the distance are refer to VQE of the most shortest distance. And then, act the action learn using that of Q value. According to the Voronoi division, the overlapping area within the range of VQE is divided finely. In figure 3, the old areas are used as Voronoi division. The red circles are created as VQE when the input enters in such a place.

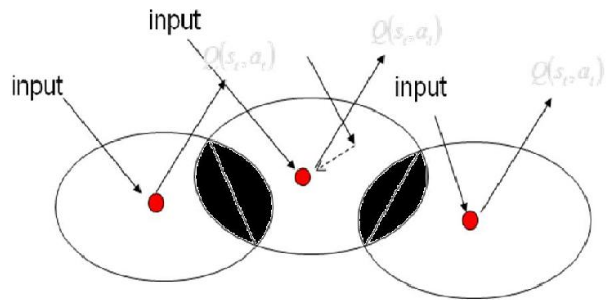

Fig3. Shape of VQE

C. Division Method of VQE

The division method of VQE is arranged by VQE in mostly at the place that a lot of points of VQE.

V. EXPERIMENT

1. 4-D Model of VQE

Figure 4 is the shape of continuous state space of VQE in 4 dimensional spaces. The content of simulation is as follows: the control action of agent is three types of forward, left rotation and right rotation. The input states of four values are the distance from the agent to the reward area1, reward area2 and the angles between the agent and those two areas.

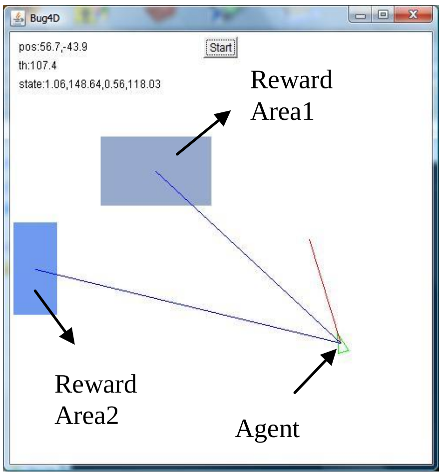

Fig.4. 4-D Model of VQE

2. Comparison of Q table and VQE

Figure 5 shows the number of rewards on the amount of one millions times over one time in four dimensional space of Q table and Voronoi Q value elements (VQE) with the constant radius when the number of state are same.

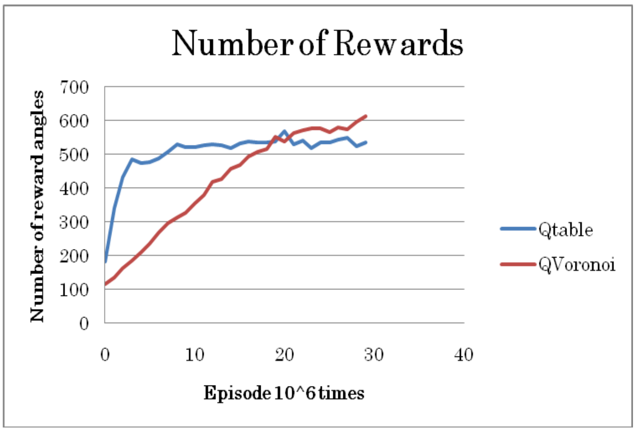

Fig.5. Q table and VQE of Reward Numbers

VI. CONCLUSION

By adopting the concept of the Voronoi diagram, VQE is constructed when the inputs are entered partially in the entire space. This method effectively simulates four dimensional spaces. Compared with the previous method on Q table and VQE, VQE is good performance. But it uses much more memory time and execution time than Q table. In order to increase the computational speed, we used the MD tree method. It is noted that, for with the constant radius of VQE.

Given the result of this study, it has been acknowledged that the VQE was suitable in higher dimension for the space division of RL with continuous state space and it was able to be confirmed that the number of elements was related to convergence time of study and a final amount of acquisition reward in VQE.

Advantages of Voronoi Space Division

It has three points. Firstly, if we use the rectangular Q elements (QE), they have the overlapping area, but when we use the Voronoi division, it does not have the unnaturalness of overlapping within the range of VQE. Secondly, when data are deleted, other data can supplement the space in the erased part. The last point is, it becomes easy to integrate.

FURTHER STUDIES

Firstly, it is necessary to adjust the number of elements of VQE. It is because the influence is whether to reach standing up and the optimum solution for study by the number of VQE. So I intend to make without the radius of VQE and VQE of movement, division and integration.

Secondly, needs to find another best way to decide which VQE is closed to the input vector by using Delaunay triangulation.

Finally, will corresponds the executed of four dimensions to a high level dimensions.

REFERENCES

[1] Richard S. Sutton and Andrew .G. Barto, Reinfo rcement Learning: An Introduction Cambridge, MA: MIT Press, 1998.
[2] Nobuyuki TANIGUCHI, Chuichi ARAKAWA and Toshio, KOBAYASHI, Construction of a Flow-Simulating Method with Finite Volume Based on a Voronoi Diagram.
[3] Peter Dayan, Christopher JCH Watkins, Reinforcement Learning
[4] Yuta HOKAMURO and Takayasu FUCHIDA, Research of Reinforcement Learning in Continuous State Space (in Japanese), 2008, Kyushu Conference, IEICE (The Institute of Electronic, Information and Communication Engineers), 09-1A-10.

