Switching Q-Learning Using Golden Cross on Hunter Games

Taisuke Ikimi  Non-member  (Osaka Prefecture University)
Keinosuke Matsumoto  Member  (Osaka Prefecture University)
Naoki Mori  Non-member  (Osaka Prefecture University)

Keywords  : machine learning, Q-learning, switching algorithms, golden cross, agents, hunter game

Recently, applications of multiagent systems are expected from the viewpoint of the parallel and distributed processing of systems. Reinforcement learning attracts attention as an implementing method of multiagent systems. However, there is a problem that the more the number of agents to deal with increases, the slower the speed of learning becomes. To solve this problem, we propose a new reinforcement learning method that can learn quickly and reduce amount of memory. It tries to increase efficiency of the learning on a hunter game by paying attention to partial states of two agents among a large number of agents. In addition, we propose a way to solve the dependability of parameters that determine a switching time of learning methods used by the existing technique, Hybrid Switchable Centralized Modular Q-learning.

On this research, we adapt Q-learning to hunter games. A unit time that each agent takes one action is called a time step, and a period of time from an initial state to a goal one (i.e. hunters catch a prey) is called an episode. If size of fields is $m \times m$ and the number of agents is $n$ of hunter games, then an agent is able to recognize $m^2n$ states. Because the exponent includes $n$, an explosion of the number of states is unavoidable. In order to avoid the problem that the speed of learning becomes slower as the number of agents to deal with, Ono and Fukumoto proposed Modular Q-learning (MQL) in 1996. MQL deals with partial states, which include two agents, therefore the number of states is reduced to $m^4$. MQL can avoid explosion of the number of states, but when agents exist, MQL needs $n(n-1)$ learning tables, so a lot of memories are needed, and the precision of learning is reduced because of an incomplete perception.

To raise the precision, we propose a new reinforcement learning method. It deals with partial agents like MQL, but there are two improved points: First, an agent has only one learning table, therefore the number of learning tables is reduced. This learning method is called “Centralized Modular Q-learning (CMQL)”. Second, the proposed method employs a switching algorithm and detects automatically a switching time by using a special index called golden cross. In early learning stage, agents learning action sequences by CMQL, and in later learning stage, agents switch learning method CMQL into “Completely Perceptual Q-learning (CPQL)”. The golden cross is a major sign for stock price analysis, and applying the golden cross to CMQL to find a switching time that learning accuracy deteriorates and the number of steps to catch a prey increases. Architecture of the proposed method is shown in Fig. 1, and a concrete algorithm for applying the golden cross to CMQL is shown in the following:

1. Set episodes of long-term average to $E_l$ and episodes of short-term average to $E_s$.
2. Calculate an average value $\bar{x}_l$ of time steps to catch a prey for $E_l$ from the present.
3. Calculate an average value $\bar{x}_s$ of time steps to catch a prey for $E_s$ from the present.
4. Switch if $\bar{x}_s$ is bigger than or equal to $\bar{x}_l$.

We tried computational experiments on hunter games with 3 hunter agents and set $E_l=900$ and $E_s=225$ episodes for example. As the result of the experiments, we observed that the proposed method is an efficient reinforcement learning method.

---

**Fig. 1. Architecture of the proposed method.**

**Fig. 2. Results of the proposed method and CPQL.**
Future of Accelerator Based BNCT Neutron Irradiation System using Liquid Lithium Target

Tooru Kobayashi Non-member (Kyoto University, kobato@rrl.kyoto-u.ac.jp)
Noriyosu Hayashizaki Non-member (Tokyo Institute of Technology)
Tatsuya Katabuchi Non-member (Tokyo Institute of Technology)
Tetsuya Yamamoto Non-member (University of Tsukuba)
Kuniaki Miura Member (Sukegawa Electric Co. Ltd.)
Masanori Aritomi Non-member (Tokyo Institute of Technology)

Keywords: BNCT, RFQ proton accelerator, liquid lithium, neutron producing target, near threshold reaction

The neutron irradiation system (NIS) for boron-neutron capture therapy (BNCT) can supply low-energy neutron irradiation field less than several tens of keV. The NIS for BNCT (BNCT-NIS) has two kinds, one is the Reactor based BNCT-NIS for neutrons from research reactors, and the other is the Accelerator based (Acc-based) BNCT-NIS. All nuclear reactors have similar characteristics of its neutron energy spectra because they are produced by the $^{235}\text{U}$ fission reaction during the critical state. They are capable of generating superior neutron intensity and temporal stability. Recently, progress in the Acc-based BNCT-NIS development has been achieved, because the improvement of the accelerator technology has overcome the neutron intensity shortage. The $^{7}\text{Li}(p,n)^{7}\text{Be}$, $^{9}\text{Be}(p,n)^{9}\text{B}$, $^{9}\text{Be}(p,xn)X$ reactions have become the viable candidates for neutron production. The characteristics of the produced neutrons, such as energy spectra and its angular dependence, varies for each reaction. Moreover, Acc-based BNCT-NIS needs to satisfy not only the necessary conditions for BNCT irradiation, but also the stability required of a neutron source.

Technical main issues related to Acc-based BNCT-NIS are the heat removal of 30-80 kW and for the radiation damage at the neutron producing target. Accordingly, high reliability which is brought about by the safety, stability, security of the system, is required as a clinical implementation. Based on the viewpoint of security of BNCT, the design standards for the Acc-based BNCT-NIS development was suggested by way of considering the secondary cancer induced by BNCT both in the affected and unaffected areas.

When the research and development of Acc-based BNCT-NIS approaches its practical implementation, the following important matters should be emphasized. The conditions and requirements related to BNCT should be carefully evaluated as much as possible. Scientists and engineers should find not only a solution and/or reduction of faults and/or the risk, but also should inspect these characteristics scientifically with a positive point of view.

The novelty of this article are the following: (1) A new design standard for an Acc-based BNCT-NIS derived from the viewpoints of secondary cancer induced by BNCT was proposed. (2) Cyclotron-Based Epi-thermal Neutron Source (C-BENS) which was realized by the joint project of Kyoto University and Sumitomo heavy industry was analyzed and evaluated using a proposed guideline. (3) The accelerator BNCT-NIS using neutrons from $^{7}\text{Li}(p,n)^{7}\text{Be}$ near threshold reaction was evaluated as a good combination with the liquid lithium target. A stable liquid lithium film flow was established to use for the neutron producing target (cf. Fig. 1).

Because of its capability for selective treatment of only tumor cells, BNCT can contribute to the improvement of the field of radiotherapy by enabling the complementary relationship with other radiation therapy modality. And BNCT has the potential for providing patient-tailored cancer therapy by combining surgery, chemotherapy, immunotherapy as necessary.

It was also found that $^{7}\text{Li}(p,n)^{7}\text{Be}$ near threshold reaction using a liquid lithium target is the most suitable neutron source for an Acc-based BNCT-NIS if it is to be realized in a hospital-setting at a downtown hospital. The next generation of BNCT irradiation systems is envisioned to have a non-invasive dose monitoring system during clinical BNCT. The combination of a long-life neutron producing target such as a liquid lithium target and a stable proton accelerator such as Radio Frequency Quadrupole (RFQ) is evaluated as a promising candidate for this future system (cf. Fig. 2).