Transactions of the Korean Nuclear Society Autumn Meeting
Goyang, Korea, October 24-25, 2019

The feasibility study of the self-controlled nuclear power plant using reinforcement learning

Jang Chansu *, Sun Jungwon, Nam Younduk, Lee Yongshin, Song Younghan, Yun Hangu, Choi Kyoungsik
Korea Nuclear Fuel, 242, Daedeok-daero 989 beon-gil, Daejeon, 34057, Republic of Korea
Corresponding author: csjang@knfc.co.kr

1. Introduction

Recently, the artificial intelligence (AI) is greatly developed and expanded into the various kinds of industries. In the nuclear industry, many researchers have tried to employ the artificial intelligence to predict the parameters in the nuclear power plant[1,2]. Most of these works was done based on the supervised learning method. To adopt the supervised learning, the huge amounts of data are required to train and validate the artificial intelligence model. The characteristics of nuclear industry does not make an access to those big amount of data. Also data cannot be generated on purpose because it can threaten the safety of nuclear power plant. These reasons introduce to employ the reinforcement learning model[3]. In this study, the reinforcement learning model is introduced to operate the virtual nuclear power plant by itself. Because there is no way to operate the real nuclear power plant to train the artificial intelligence, the virtual environment (nuclear power plant) was constructed using PYTHON computer language. The virtual environment is connected to the reinforcement models, i.e., Deep Q Network (DQN) and actor-critic models. The objectives of this study are to investigate the possibilities of the self-controlled nuclear power plant using the reinforcement learning method.

2. Methods and Results

In this section the reinforcement learning model constructed in this study are described.

2.1 Building of Virtual Environment

Due to the impossible operation of real nuclear power plant, the building of virtual nuclear power plant is required to train the artificial intelligence. Fortunately, the safety analysis computer code can simulate the every operational component and the calculation methods of the code have licensed by the nuclear regulatory body. It means to be a safe and sound method to show that the nuclear power plant can be operated safely by the artificial intelligence. The safety analysis code was made by FORTRAN computer language at 1960s and the variety of modifications must be performed to use in the reinforcement learning method. Because the PYTHON computer language is most powerful to implement the artificial intelligence model, the safety analysis code was dismantled and reassembled to build the PYTHON extension module. The f2py tool was the best way to make the PYTHON extension module from FORTRAN source code. The purpose of the f2py (FORTRAN to PYTHON interface generator) is to provide a connection between PYTHON and FORTRAN languages. gfortran and gcc compilers were used to compile the remodeled safety analysis code. Fig. 1 shows the comparison results of the original code (FORTRAN) and remodeled code (PYTHON) calculation. Two lines are on each other.

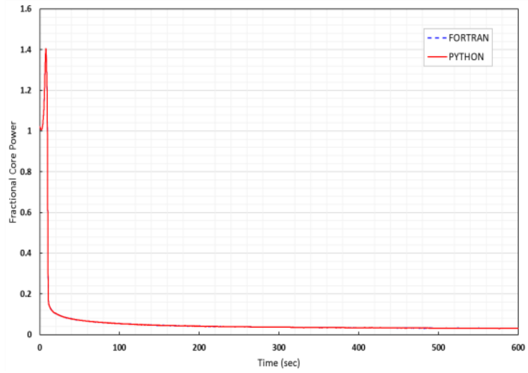

Fig. 1. Comparison of FORTRAN and PYTHON safety analysis codes

2.2 Agent Models

Reinforcement learning [3] is learning what to do and how to map situations to actions so as to maximize a numerical reward signal.

To train the artificial intelligence, the agent models must be designed and constructed. There are many methods of reinforcement learning for constructing agent model. In this study two methods are used i.e. Deep Q Network (DQN)[4] and Actor Critic[5]. DQN is very famous with winning the human in the Go game at 2016. The idea of DQN is to create a neural network that will approximate, given a state, the different Q-values for each action. Actor-critic methods aim at combining the strong points of actor-only and critic-only methods. To study the sensitivity depending on the agent models, the above two models were used to create the agent.

2.3 Selection of Transient Accident

Many kinds of accidents can be occurred in the nuclear power plant by the various initiating events. To train the artificial intelligence, an adequate transient accident must be chosen. The chosen accident must not be too much complex and be enough good to show the nuclear power plant operating capability of the artificial intelligence. If the transient is too complex, the artificial intelligence is not sure to control the situation. And the project can be failed without any meaningful outcomes.

In this study, the single control element assembly withdrawal (SCEAW) accident is selected to search for the possibility of the self-controlled nuclear power plant. The SCEAW accident is initiated by the withdrawal of an inserted control rod out of the reactor core. The core power increases due to the positive reactivity addition by SCEAW. This accident is relatively slow transient compared with the other reactivity-induced accidents. Because of this characteristics, the artificial intelligence can follow and control the situation successfully. Fig. 2 shows the trend of core power with time during SCEAW accident.

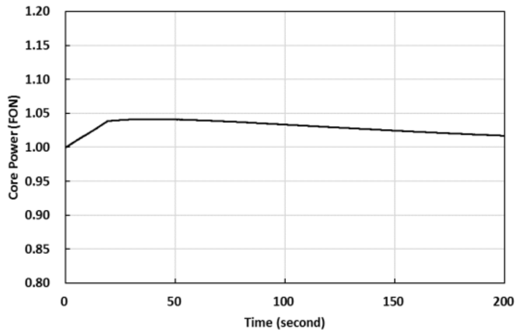

Fig. 2. Core power vs. Time during SCEAW accident.

2.4 Design of State, Action and Reward

To adopt the reinforcement learning method, the state and action must be defined adequately. If the state and action are defined in the wrong direction, the reinforcement learning would be failed. Because the purpose of this study is to find the possibility of self-controlled nuclear power plant, the important parameter will be the core power. Therefore the core power will be a state and the rod maneuvering control will be actions. The reward is 1 point if the core power is controlled within a target range and 0 point if the core power is controlled out of a target range. Because the total number of steps is 1709, the total reward which the artificial intelligence can get will be 1709 points. If it gets the full score, it will be considered that the nuclear power plant is controlled successfully by the artificial intelligence.

2.5 Experimental Train Results

After combining the reinforcement learning method and the virtual nuclear power plant, the various kinds of train with the different hyper parameters (learning rates, number of actions, number of states, and discount factors) were performed. Many train cases were failed to converge and some of cases were successful. The greatest train case have reached over the full score. Fig. 3 shows the comparison results of the artificial controlled core power with DQN model (AI) and the original accident core power (No AI) during the SCEAW accident described in Section 2.3. When the artificial intelligence does not get involved in the plant operation, the core power increases about 5% due to the single control element assembly withdrawal. But if the artificial intelligence starts to control the nuclear power plant, there is no power increase. The core power keeps constant around the nominal power during the entire range of transient.

Even if this breakthrough results can be shown, the reinforcement learning method does not show the reliable and constant performance after the enormous train episodes. Especially actor-critic method seemed not to reach the convergence. Generally actor-critic method is converged faster than DQN method. In this study DQN is a better artificial intelligence method to control the nuclear power plant.

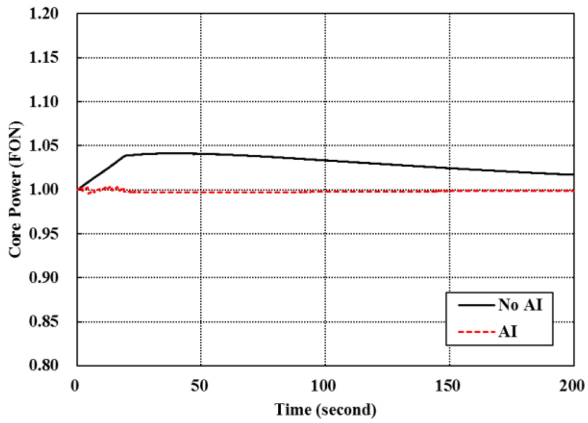

Fig. 3. Comparison of AI-controlled core power and accident core power during SCEAW accident

3. Conclusions

To investigate the feasibility of self-controlled nuclear power plant based on the reinforcement learning method, the virtual nuclear power plant was constructed and then linked with the artificial intelligence models. The adequate states, actions and reward system were designed for the train of artificial intelligence. The slow reactivity-induced accident was selected to train the artificial intelligence. DQN and actor-critic method were used to control the nuclear power plant. DQN method shows a magnificent results at some cases but actor-critic method seemed not to be converged.

Through this study, the possibility of self-controlled nuclear power plant was shown. The reliable and constant performance is still remained as a problem to be solved.

REFERENCES

[1] Kyung Woo Shim and Dong Su Kim, "Axial Power Distribution Synthesis for In-COre Protection System (ICOPS) Using Artificial Neural Network with a Single Hidden Layer", KNS Spring, 2019.
[2] K.W. Shim, D.Y. Oh, D.S. Kim, Y.J. Choi, and Y.K. Park, "COLSS Axial Power Distribution Synthesis Using Artificial Neural Network with Simulated Annealing", KNS Spring, 2015.

[3] Richard S. Sutton and Andrew G. Barto, "Reinforcement Learning – An Introduction", 2nd Edition, The MIT Press.
[4] Volodymyr Mnih, etc, "Human-level control through deep reinforcement learning", Nature, 26th February 2015
[5] Vijay R. Konda and John N. Tsitsiklis, "Actor-Critic Algorithms", Laboratory for Information and Decision Systems, Massachusetts Institute of Technology