Optimization of binary decision diagram: Heuristics from reinforcement learning

Young Ho Chae [a*], Poong Hyun Seong [a]
[a] *Department of Nuclear and Quantum Engineering, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea*
[*] *Corresponding author: cyhproto@kaist.ac.kr*

1. Introduction

Since nuclear power plants (NPPs) are safety critical infrastructure, enhancing safety of nuclear power plant is critical. To evaluate the degree of safety probabilistic safety assessment method is widely used. And to assess the probability, various techniques were invented. Fault tree (FT), reliability block diagram with general gate (RBDGG), universal generating function (UGF), and binary decision diagrams (BDD) are typical examples. However, none of the method is perfect for all circumstance. Each analysis method has their own pros and cons. For instance, fault tree has benefits when identifying the logical fault from effect. However, as system becomes complex, the tree structure becomes even more complex. RBDGG is firstly proposed by Kim [1]. RBDGG method is intuitive however, the method is hard to consider the system in logical loop and minimal cutsets. To solve the logical loop problem, Chae et. al. [2] suggested factor graph based approach. But the method is only suitable for the logical loop structure. Universal generating function which was developed by G. Levtin [3] has benefits when analyzing multi-state systems and degradation of the system. However, UGF requires prior definitions of various arithmetic functions. Binary decision diagram is the most intuitive method among them. However, the method has critical deficiency which is called as variable ordering problem. The complexity of the diagram depends on the order of the variable. And finding the optimal order of the variable is typical example of non-deterministic polynomial (NP) problem.

In this paper, we propose reinforcement learning based variable ordering to find approximate optimal solution of the variable order of binary decision diagram. The organization of the paper is as follows. In section 2, a brief explanation about reinforcement learning will be provided. In section 3, general structure of the agent and reward function which is the most critical part in reinforcement learning will be discussed. Finally, section 4 presents conclusion.

2. Reinforcement learning

The machine learning (ML) algorithms can be sorted by the learning method. The first type of ML is supervised learning. The objective of supervised learning is finding the mapping function with pre-defined label. Supervised learning is suitable for regression and classification problem. On the contrary, unsupervised learning utilizes data without pre-defined label. The objective of unsupervised learning is finding the cluster. The last algorithm is reinforcement learning. The objective of reinforcement learning is finding the proper strategy in a given environment by using environment agent interaction. The early form of reinforcement learning algorithm is introduced by Sutton and Barto [4].

The other two methods use existing data, but reinforcement learning is a method of finding strategies through interaction with the environment rather than using existing data. Due to this perspective, the algorithm has different characteristics from the two methods.

The reinforcement learning agent utilize two strategies during seeking the goal. The first strategy is exploration. An agent must prefer actions that it has tried in the past and found to be effective in producing reward (Exploration). Also, an agent has to exploit what it already knows in order to obtain reward, but it also has to explore in order to make better action (Exploitation).

Reinforcement learning is composed with policy, reward signal, value function, and model of the environment. The policy is the probability of action a when the agent is in state s. Mathematically, the policy function (π) can be defined as following equation.

$$\pi(a|s) = P(a|s)\ (Eq.1)$$

The reward can be defined in various ways depending on the situation, but it is generally defined as the expected value under situation S. The accumulated reward function can be written as follow (Eq.2).

$$R(\boldsymbol{S}) = r(s_0) + \gamma r(s_1) + \gamma^2 r(s_2) \ldots (Eq.2)$$

Although the present gain is important, the expected rewards in the future must also be considered, therefore to calculate the accumulated reward, the future reward is also added. However, a discount rate is applied to the future value in order to discount the future reward.

The value function is expected reward by using policy Q at state S. Value function can be written as follow (Eq.3)

$$V^{\pi}(s) = E[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s, \pi]\ (Eq.3)$$

Using the above elements, reinforcement learning is performed in the direction of determining the next action to be performed through the policy function, evaluating the result of the action using the value function, and

improving the policy. The basic sequence and elements of reinforcement learning are as above. Deep reinforcement learning is a model that uses deep learning in designing and improving policy tables in the process.

3. Reinforcement learning agent for binary decision diagram

To design reinforcement learning model, state, action, and reward should be defined. In this paper we defined elements as follows.

3.1 State

A state s which is a subset of state vector S is defined as the form of the partially constructed BDD with variable order. For instance, let assume that there are four variables which is denoted as X1, X2, X3, and X4. And the agent firstly selected X2 and secondly selected X3. Then in this case, the state is the form of partially constructed BDD with variable order X2, and X3.

3.2 Action

An action a which is a subset of possible action set A is defined as the selecting variable at state s.

3.3 Reward

For the reinforcement learning agent, reward is critical factor because, the direction of learning can be different depending on the design of reward function. We designed the reward based on information entropy. The concept of information entropy is firstly introduced by Shannon [5]. By using information entropy, the impurity of set can be defined. We designed reward as eq. 4.

$$\mathrm{R(S)} = \sum_{k=0}^{end\ of\ the\ state} \frac{\gamma^k}{H(S_k)} \quad Eq.4$$

For instance, let assume there are two different agent and both of them selected x1 as first variable. However, in second action, agent 1 selected x2 as second variable, and agent 2 selected x3 as second variable. Then one of the possible diagram structure can be drawn as follows.

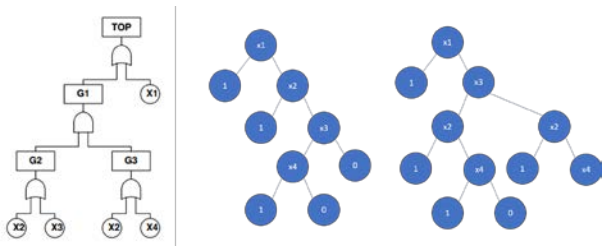

Figure 1. Fault tree model and binary decision diagram with different order

In this case, the entropy of the first figure can be calculated as 1.922 and the second figure can be calculated as 3.084. Therefore the expected rewards are 0.520 and 0.324 respectively. The agent 1's reward is bigger than agent 2. Therefore, the agent will be trained as of agent 1.

3. Conclusion

To enhance the safety of nuclear power plants, probabilistic safety assessment is crucial. For the intuitive and accurate probability calculation, several methods were invented. However, none of the methods are perfect for the all circumstance.

Binary decision diagram has strength in intuitiveness. However, the method has critical deficiency which is called as variable ordering problem. And the variable ordering problem is non-polynomial problem. Therefore the problem heavily depends on heuristics.

In this paper, to simplify the form of the BDD structure, reinforcement learning method is applied. The state is defined as partially constructed diagram structure and the action is defined as selecting variable. And the reward which is the most critical part in reinforcment learning is defined by using information entropy. With suggested concept, binary decision diagram can be utilized for the PSA.

ACKNOWLEDGEMENT

This research was supported by the National R&D Program through the National Research Foundation of Korea (NRF) funded by the Korean Government. (MSIP: Ministry of Science, ICT and Future Planning) (No. NRF-2016R1A5A1013919)

REFERENCES

[1] M.C. Kim, "Reliability block diagram with general gates and its application to system reliability analysis", Ann Nucl Energy, 38 (2011), pp. 2456-2461
[2] Young Ho Chae, Seung Geun Kim, Poong Hyun Seong, "Reliability of the system with loops: Factor graph based approach", Reliability Engineering & System Safety, Volume 208 (2021)
[3] G. Levitin, “The universal generating function in reliability analysis and optimization” 10.1007/1-84628-245-4 (2005).
[4] Sutton, R. S., and Barto, A. G., "Reinforcement learning: an introduction", Vol.1, MIT press Cambridge (1998)
[5] C.E. Shannon, "A mathematical theory of communication", Vol. 27, pp. 379-423, 623-656, July, October, 1948