Study on the effective training data for a classification model to evaluate the reliability of a passive safety system

Kyungho Jin[1], Hyeonmin Kim[1], Seunghyoung Ryu[2], Seunggeun Kim[2], Jinkyun Park[1*]

[1]*Risk Assessment and Management Research Team/* [2]*Applied Artificial Intelligence Lab,*
Korea Atomic Energy Research Institute, 111, Daedeok-daero 989, Daejeon, Republic of Korea

*Corresponding author: *kshpjk@kaeri.re.kr*

1. Introduction

Passive Safety Systems (PSSs) in Nuclear Power Plants (NPPs) are designed to perform the intended functions by natural forces such as gravity to minimize the external effects on the system (e.g., human interactions/external power) [1]. Although these inactive safety systems significantly contribute to the safety of NPPs, it is difficult to evaluate the reliability of them due to the lack of data, insufficient understanding of the phenomenon and not enough operating experiences.

In order to estimate the reliability of a PSS, a framework that employs Monte Carlo (MC) simulation with a thermal-hydraulic (T-H) code has been proposed [2][3]. The T-H code (describing the target PSS) derives the output results by propagating the uncertainty of input variables using MC simulation. The success or failure of the system is then classified by comparing to a failure criterion. As a result, the reliability of the PSS can be determined through the number of the failure cases.

However, this approach requires the considerable efforts (e.g., computing power) because MC simulation requires a large amount of T-H code runs. Recently, with the growth of deep learning technology, the efforts are being made to reduce the cost of analysis by developing a surrogate model (e.g., classification model) that replace the T-H code [1][4][5]. There is no significant difference from the existing MC simulation except that a surrogate model instead of the T-H code produces the results.

Unfortunately, there are still several factors that make it difficult to evaluate the reliability of a PSS through a surrogate model-based approach; the training data can be obtained only from a T-H code and a relatively large amount of training data is still required to fit hyper-parameters of a surrogate model. In addition, the generated data are extremely imbalanced. While most of the cases describe the success of the system, it rarely indicates the failure due to the high reliability of a PSS.

Therefore, in this paper, the effective method for classification has been proposed to estimate the reliability of a PSS with reduced computational burdens by dealing with the data imbalance during training data generation. Section 2 introduced the general procedure to develop a classification model. Section 3 explained how to deal with data imbalance and Section 4 showed the results of the case study.

2. A surrogate model-based reliability evaluation of a passive safety system

2.1. MC simulation with a T-H code

Using MC simulation with a T-H code, the reliability of a PSS can be evaluated through Eq. (1) [1].

$$P_f = \int \dots \int I_g(X_d) f(X_d) dX_d \qquad (1)$$

where, P_f is a failure probability of a PSS and X_d is a set of input variables used in a T-H code ($X_d = [x_1, x_2, \dots, x_d]$). If a T-H code and output of T-H code is denoted by Y_{X_d} and $g(X_d)$ respectively, $I_g(X_d)$ is an indication function (e.g., if $g(X_d) > c_f$ then 1, otherwise 0; c_f is a failure criterion). $f(X_d)$ is a joint probability function of X_d. According to the law of large numbers, Eq. (1) can be approximated as the total number of failures (N_{fail}) divided by the total number of simulations (*N*).

In order to estimate the reliability of a PSS through Eq. (1), X_d and Y_{X_d} should be simulated. Generally, X_d are assumed to be independent, therefore, they are generated from $f(X_d) = \prod_{j=1}^{d} f_j(x_j)$ (*j* = 1 to *d*). When $X_d^1, \dots, X_d^N$ are organized, the outputs (Y_{X_d}) are derived using $g(X_d^k)$. The success/failure of the system is determined by comparing to c_f. Consequently, $\widehat{P_f}$ can be evaluated by $\frac{N_{fail}}{N}$.

2.2. Development of a surrogate model

Over the past decade, there have been great advances in deep learning technology. In the nuclear industry, many research areas are also employing the concept of deep learning. One of the applications is to develop a model to replace the T-H code for fast simulation. This fast simulation approach has also been implemented to the reliability estimation of a PSS [1][4][5].

Although there are various surrogate models, this paper focused on a classification model. What is different from the existing method (e.g. described in Sec. 2.1.) by developing a surrogate model is that X_d and Y_{X_d} is not directly used to estimate the reliability of a PSS, but utilized to fit a surrogate model. In other words, X_d and Y_{X_d} becomes training data.

This paper employed the simple Fully Connected Neural Network (FCNN) for classification. Since the designing and training of a classification model using neural network is well known, the details were omitted in this paper.

When a surrogate classification model denoted by $\hat{g}(X_d)$ is established, the reliability of a PSS can be evaluated in the same way as in Sec. 2.1. The overall procedure was described in Table I.

Table I. The overall procedure of a surrogate model-based reliability evaluation of a PSS

Step	Procedures
1	**Simulate** X'^k_d from $f(X_d)$ (k = 1 to N)
2	**Derive** $Y'^k_{X_d}$ from T-H code $g(X'^k_d)$
3	**Train** a surrogate model $\hat{g}(X'_d)$ using X'^k_d and $Y'^k_{X_d}$
4	**Simulate** X^k_d from $f(X_d)$ (k = 1 to N)
5	**Derive** Y_{X_d} from the surrogate model, $\hat{g}(X_d)$
6	**Compare** Y_{X_d} to c_f (1 = *failure*, 0 = *success*)
7	**Calculate** $\widehat{P_f} = \frac{N_{fail}}{N}$.

Note that the descriptions of step 1 to 3 in Table I belongs to Sec. 2.1 and step 4 to 7 belongs to Sec. 2.2. It should be noted that X'^k_d is totally different from X^k_d. X^k_d was used to estimate the reliability of a PSS whereas X'^k_d was used to fit a surrogate model.

3. The effective training data for a classification model

3.1. Imbalance in training data

Generally, PSSs have a high reliability because external factors that cause the system failure are excluded. This characteristic of a PSS is obviously helpful to enhance the safety, but it makes it difficult to estimate the reliability of them using a surrogate model.

For example, let us suppose that the failure probability of a certain PSS is about 1.0E-04 and we have 10,000 simulation results of X'^k_d and $Y'^k_{X_d}$, respectively. $Y'^k_{X_d}$ is then classified as *success* or *failure* by comparing to c_f for training a classification model. In this case, most cases include *success* (e.g., about 9,999), but only few cases (e.g., about 1) indicate *failure*. Fig. 1. shows an example of 1,000 imbalanced data when a c_f is assumed to be located in the tail of the output distribution.

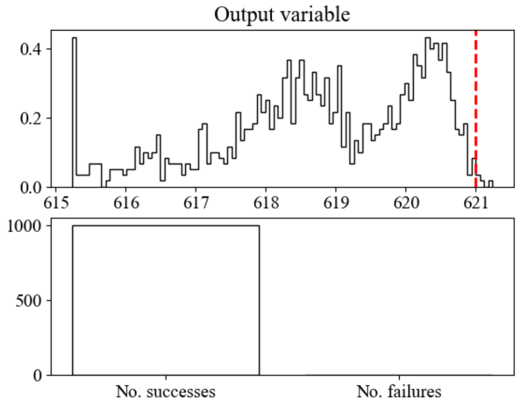

Fig. 1. An example of imbalanced data when a c_f is assumed to be located in the tail of the output distribution

When these data are used as training data, the performance of a classifier that has not experienced much failure will be apparently not good. Data imbalance eventually results in an increase in computational burdens because more failure cases are required to guarantee the surrogate model performance; it counteracts the advantages of employing a surrogate model.

3.2. How to obtain the effective training data for classification

One simple way of resolving the above problem is to adjust the probability density function $f(X_d)$ of input variables so that the output variable includes more failures. Unfortunately, it is not easy to identify the adjusted probability distribution considering this information.

Therefore, this paper tried to deal with the data imbalance problem by dividing it into the initial data generation part and the additional data generation part. In the former part, X'^m_d are generated from $f(X_d)$ as presented in Table I. In the latter part, X'^v_d are generated from the Empirical Cumulative Density Function (ECDF) instead of $f(X_d)$.

The reason why the ECDF was used in this paper is that it is uncomplicated to consider the failure information of the output to the uncertainty distribution of the input variable. For example, if X'^m_d and $Y'^m_{X_d}$ are constructed in the former part, we can easily obtain the cases of input variables that cause output variable to fail. The ECDFs of each input variable are then built based on this failure information. Now, sampling the input variable from these ECDFs can make the output contain more failure cases. The concept of sampling from ECDFs are shown in Fig. 2.

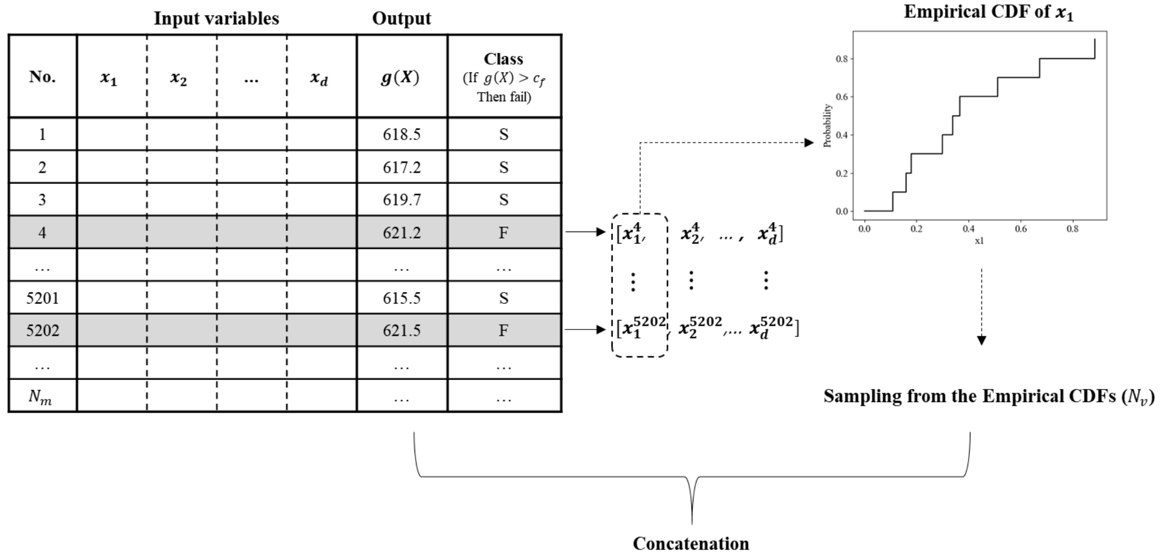

Fig.2. The concept of sampling from the ECDFs based on the failure cases

Finally, the training data is merged by $(X'^{m}_{d}, Y'^{m}_{X_d})$ and $(X'^{v}_{d}, Y'^{v}_{X_d})$. In this way, more failure information can be achieved using the same number of training data. For example, Fig. 3. shows the 1,000 simulation results using the proposed method compared to Fig. 2.

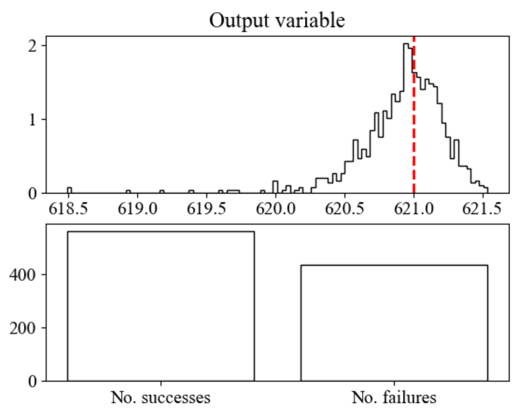

Fig. 3. Additional training data with more failure cases

If the data imbalance is resolved well, a classifier with higher performance can be built with the same number of training data. This means that the computational burden can be reduced.

4. Case study

In the case study, the estimation results of P_f depending on the generation methods of training data were illustrated. Advanced Thermal-hydraulic Test Loop for Accident Simulation (ATLAS) in Korea Atomic Energy Research Institute (KAERI) was used as a target PSS in the case study. Multi-dimensional Analysis of Reactor Safety (MARS) was used to model the target PSS.

Total 15 input variables and 1 output variable were identified. In addition, 100,000 cases were simulated to compare the results. The failure criterion, c_f was assumed to be 621. As a result, total 586 cases were identified as a failure; the reference failure probability of the target PSS was 5.86E-03.

The classification model based on FCNN used in the case study had 4 layers (each layer has 50, 500, 500, 50 nodes, respectively). 30% of the total data were set to test data and 70% were used as training data. 10% of training data were used to validate the classification model.

Fig 4. shows the estimation results of P_f based on the original generation method of training data (e.g., described in Sec. 2). 3 repetitions were taken to account for the uncertainty of sampling. As the total number of data increases, it approached the answer, but it can be seen that the uncertainty is high in the case where a small number of data were used.

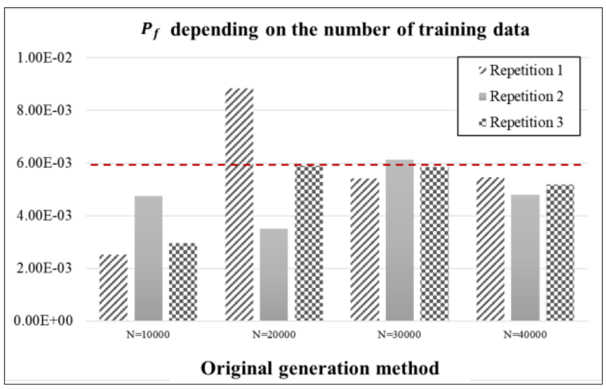

Fig. 4. the estimation results of P_f using the original method for the generation of the training data

On the other hand, Fig. 5. shows the estimation results of P_f using the proposed method (described in Sec. 3). 10% of training data were additionally generated in the case study (e.g., Initial generation: 9,000 + additional generation: 1,000 = total 10,000 training data were used). Compared with Fig. 4, it can be seen that the estimation results are much closer to the answer even with a smaller number of data. In addition, there were less uncertainty with respect to sampling.

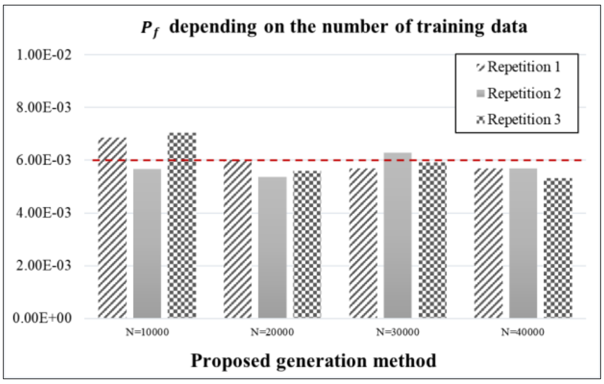

Fig. 5. the estimation results of P_f using the proposed method for the generation of the training data

Consequently, it was confirmed that it is possible to build a high-accuracy classifier with a smaller number of data through the proposed method. It will lead to a reduction in computational burden in the reliability evaluation of a PSS.

5. Conclusions

This paper proposed the generation of effective training data for a classification model to reduce computational burdens in the reliability evaluation of a PSS. The proposed model was composed of the initial generation and additional generation part. In the latter part, adjusted probability distributions using ECDF were constructed to deal with data imbalance by containing more failure cases.

The case study showed that a high-accuracy classifier with a smaller number of data can be established using the proposed method. However, it was limited to a classification model and how many additional data should be generated was not considered in this paper. Furthermore, more case studies should be performed to confirm the uncertainty with the configuration model.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Science, ICT, and Future Planning of the Republic of Korea and the National Research Foundation of Korea (NRF-2019M2C9A1055906, NRF-2020M2C9A1061638).

REFERENCES

[1] E. Zio and N. Pedroni, "How to effectively compute the reliability of a thermal – hydraulic nuclear passive system," *Nucl. Eng. Des.*, vol. 241, no. 1, pp. 310–327, 2011, doi: 10.1016/j.nucengdes.2010.10.029.

[2] M. Marqu, J. F. Pignatel, P. Saignes, F. D. Auria, and L. Burgazzi, "Methodology for the reliability evaluation of a passive system and its integration into a Probabilistic Safety Assessment," *Nucl. Eng. Des.*, vol. 235, pp. 2612–2631, 2005, doi: 10.1016/j.nucengdes.2005.06.008.

[3] A. K. Nayak, M. R. Gartia, A. Antony, G. Vinod, and R. K. Sinha, "Passive system reliability analysis using the APSRA methodology," vol. 238, pp. 1430–1440, 2008, doi: 10.1016/j.nucengdes.2007.11.005.

[4] N. Pedroni and E. Zio, "An Adaptive Metamodel-Based Subset Importance Sampling approach for the assessment of the functional failure probability of a thermal-hydraulic passive system," vol. 48, pp. 269–288, 2017, doi: 10.1016/j.apm.2017.04.003.

[5] H. Kim, J. Cho, J. Y. Yoon, and J. Park, "The generation of failure surface for reliability assessment of passive safety systems using deep learning technology."