Study for Machine Learning Model to Predict the Sequential Event of Severe Accident Depending on Operator Action

Dae Hyung Lee [a*], Sun Hong Yoon [a]
[a] *KEPCO E&C, Nuclear Technology Research Dept., 269 Hyeoksin-ro, Gimcheon-si, 39660*
[*] *Corresponding author: bigbro@kepco-enc.com*

1. Introduction

Predicting the occurrence time of sequential events resulting from the operation of the mitigation systems in the severe accident can help operators to make appropriate decision to mitigate the accident.

In this study, we developed a model to predict sequential events varying with the operation time of the mitigation systems after the severe accident.

The APR1400 severe accidents initiated by the Large Break Loss of Coolant Accident (LBLOCA) were simulated using MAAP version 5.03 with the failure of the safety injection system. In this scenario, it is assumed that the operator restores and activates the mitigation systems after core damage. Since the restoring system and making decision takes time, the input of operation time varies.

The machine learning (ML) model was developed to predict the occurrence time of a particular event by inputting the information on the initial event and the operation time of mitigation system instead of time serial I&C values. In particular, we focused on how to deal with the operation time of the mitigation systems and how to build a model to predict the occurrence time of sequential events- i.e., reactor vessel (RV) failure in this study.

Not utilizing the time serial data makes it simple and fast to build and train the model.

2. Methods and Results

2.1 Scenario Configuration and Data Generation

The dataset has been generated from the MAAP analysis with about 3,000 scenarios which have various break sizes, break locations such as the hot legs and the cold legs, and actuating timings of the mitigation systems after the core damage in the LBLOCA-induced-severe accidents.

The Safety Injection (SI) system, the Cavity Flooding (CF) system and the Containment Spray (CS) system are selected as the mitigation systems for the LBLOCA-induced severe accidents based on the Level 2 PSA results.

The information of MAAP input is shown in Table 1.

Table 1: Scenario (MAAP input) Configurations

Variables *		Range **
LOCA	Size (dia.)	[6~16] in
	Location	Hot/Cold leg
time period to be actuated (after core damage) ***	SIS	[900~14,400] sec
	CFS	[900~14,400] sec
	CSS	[1800~180,000] sec

* All variables are stratified by latin hyper cube sampling [1].
** The minimum actuation time is selected referring to Human Reliability Analysis.
*** 20% of all cases for each safety system is not functional to describe operation fail.

2.2 Machine Learning Model with Input of System Operation Time

The machine learning models dealing with initial accident conditions and the mitigation system operation timings as input features are much simpler than those dealing with the time serial data. In spite of their simplicity, they can give an important information to the operator if it is possible to predict the accident with sufficient accuracy.

Initial accident conditions (location/size) can be used as input features for this ML model since it has been verified through previous study [2] that they can be diagnosed with high accuracy when an accident occurs.

We can choose the actuation timings of the mitigation systems, using the Latin Hypercube Sampling (LHS) [1] within the time period defined in the Table 1.

Table 2 shows the part of dataset for the target-oriented ML model.

Table 2: Dataset for RV Failure Prediction

Input Features					Target
Break Location*	LOCA Size(in)	SI (s)	CF (s)	CS (s)	RV Failure(s)
1	14.9	4740	3505	85329	4994
0	9.0	15914	11071	27054	8983
1	10.6	10697	2680	34961	5947
0	15.0	12416	-	72269	6282
0	10.2	8760	5080	45822	8280
0	7.1	9487	12799	-	-
...					...

* 0 and 1 indicates hot leg and cold leg, respectively.

2.3 Data Preprocessing

Before training the ML models using the results from Table 2, preprocessing of null value is required.

1) Preprocessing for Not-Operated System

Operation time of a non-operated system means that it is physically later than any other operated systems, so it is not appropriate filling with zero instead of null. Because zero would be recognized earlier than the earliest operated system.

On the other hand, if the large number exceeding the analysis time range is inputted for non-operated systems, there would be a risk that the differences in system operating time that are of interest can vanish during the data scaling process.

Therefore, instead of adjusting the operating time of the system only, we solved this problem by adding columns indicating whether the systems are operational or not.

Table 3: Input Settings for system operation

LOCA Loc.	LOCA Size	SI on	SI	CF on	CF	CS on	CS
1	14.9	1	4741	1	3506	1	85330
0	9.0	1	15914	1	11071	1	27054
1	10.6	1	10697	1	2680	1	34961
0	15.0	1	12417	0	0	1	72270
0	10.2	1	8760	1	5080	1	45822
0	7.1	1	9487	1	12799	0	0
...							

2) Preprocessing for Not-occurred Target Event

In order to predict event timing, the idea proposed in this study is to add a step of determining whether the event occurs. We added a column to the target data indicating whether the event would occur and added a classification step to the Two-Step Target-Oriented (TOSTO) ML model.

When the data is inputted to the TOSTO model, the model predicts whether the event occurs or not on the first step and specific event time is predicted on the second step only if the event is expected to occur.

The schematic prediction model diagram is shown in Fig. 1.

In this study, prediction model uses random forest algorithm with default hyperparameters from the Scikit-learn 0.24.1. The random forest is an ensemble method which works well with non-linear data and has lower risk of overfitting. TOSTO model uses random forest classifier and random forest regressor in each step.

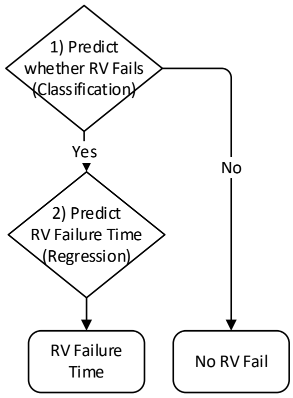


Fig. 1. Two-Step Prediction Model

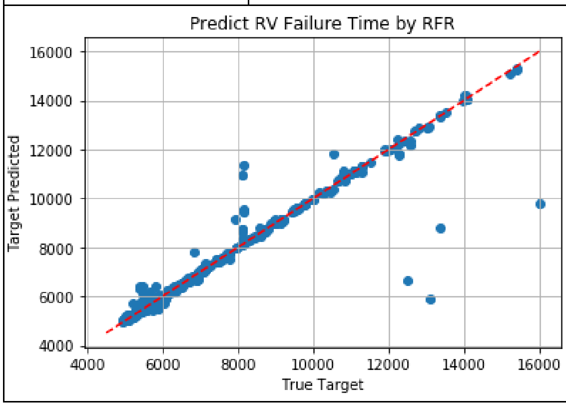
2.4 Results

The results of the two-step prediction model trained with preprocessed data are shown in Table 4. About 25% of data is used as a test dataset.

The primary classification model predicted RV failure with an accuracy of 97%, and the secondary regression model predicted RV Failure time with R^2 score of 0.93 in our TOSTO ML model.

Table 4: Two-Step Prediction Model Results

1st Step: Classification			
ML Model	Random Forest Classifier		
Accuracy	0.9724		
Confusion Matrix		Predict (N/P)	
	Target (N/P)	189	12
		8	516
2nd Step: Regression			
ML Model	Random Forest Regressor		
R^2 Score	0.9282		
MAE	137.04		



3. Conclusions

In this study, we developed a TOSTO ML model to predict the reactor vessel failure timing using initial accident condition and mitigating system operation time as input features. Methods for preprocessing non-operating system timings and separating the ML steps to predict the subsequent events are verified with high accuracy.

Since this model uses simplified data with only a few features, it is very fast in training and prediction. In addition, this model has the advantage of ignoring the uncertainty of time serial data prediction.

Although the current machine learning model shows relatively high accuracy, there is a limitation in that it is difficult to improve performance with given simple dataset. In the further study, therefore, the model performance will be improved by utilizing time serial data.

REFERENCES

[1] Daehyung Lee, Sampling Methods for Uncertainty Analysis Using MAAP5, Korean Nuclear Society, 2020
[2] Geon Pil Choi, Man Gyun Na, Estimation of LOCA Break Size Using Cascaded Fuzzy Neural Networks, Nuclear Engineering and Technology, Vol. 49, p. 495, 2017.
[3] Electric Power Research Institute, Inc., MAAP 5 User's manual, 2008.