Study for Machine Learning Model to Predict the Sequential Event

Dae Hyung Lee, Sun Hong Yoon
KEPCO E&C, Nuclear Technology Research Department, 269 Hyeoksin-ro, Gimcheon-si, 39660
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 an appropriate decision to mitigate the accident.

In this study, we developed a model to predict sequential events (reactor vessel failure) with input of the operation time of the mitigation systems after the severe accident. 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, locations such as hot leg and cold leg, and actuating timings of the mitigation systems (SIS, CFS, CSS) after the core damage in the LBLOCA-induced-severe accidents.

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
Actuation time (after core damage) ***	SIS	[900~14,400] sec
	CFS	[900~14,400] sec
	CSS	[1800~180,000] sec

* All variables are stratified by Latin Hypercube Sampling [1].
** The minimum actuation time is selected referring to Human Reliability Analysis.
*** 20% of each case is set off to depict the system failure.

2.2 Data Preprocessing

Before training the ML models, preprocessing of null value is required.

1) Preprocessing for Not-Operated System

In this section, we will discuss how to deal with the operation time of the non-operated systems. Operation time of the non-operated system is technically 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 2: Dataset for RV Failure Prediction

Input Features								Target
LOCA Loc.*	LOCA Size(in)	SIS		CFS		CSS		RV Failure Time (s)
		On	Time (s)	On	Time (s)	On	Time (s)	
1	14.9	1	4740	1	3505	1	85329	4994
0	9.0	1	15914	1	11071	1	27054	8983
1	10.6	1	10697	1	2680	1	34961	5947
0	15.0	1	12416	0	-	1	72269	6282
0	10.2	1	8760	1	5080	1	45822	8280
0	7.1	1	9487	1	12799	0	-	-
...								...

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

2) Preprocessing for Not-occurred Target Event

In order to predict the 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 ML model.

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

Two-Step ML model uses random forest classifier and random forest regressor in each step.

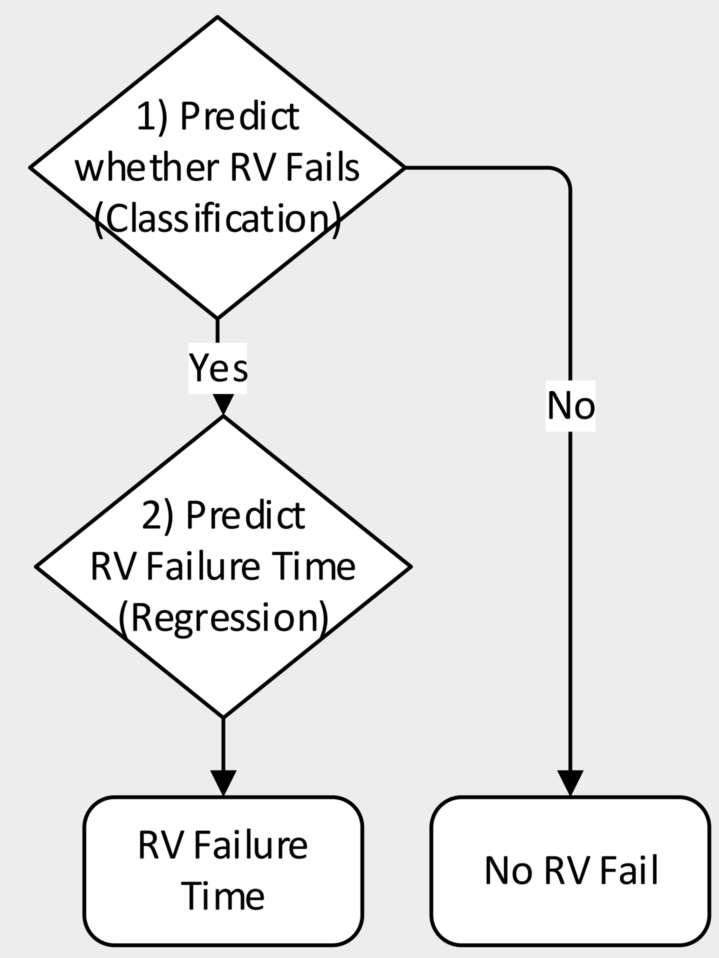
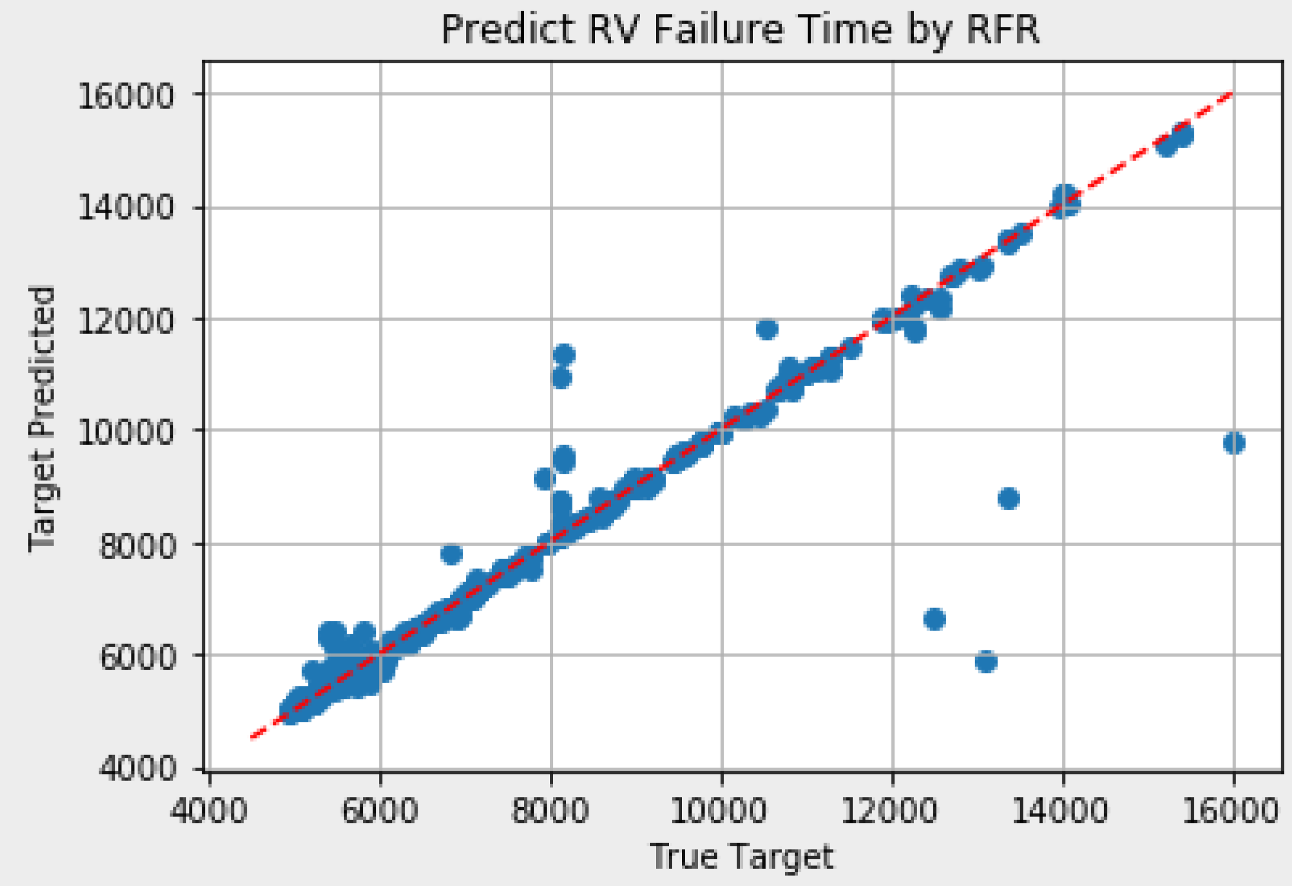


Fig. 1. Two-Step Prediction Model

2.4 Results

1st Step : Classification	
ML Algorithm	Random Forest Classifier
Accuracy	0.9724
F1 Score	0.9810
2nd Step : Regression	
ML Algorithm	Random Forest Regressor
R^2 Score	0.9282
MAE	137.04



3. Conclusions

In this study, we developed a ML model to predict the reactor vessel failure timing using initial accident condition and mitigating system operation times as input features. Methods for preprocessing non-operating system timing 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.