2021 KOREAN NUCLEAR SOCIETY

Development of machine learning methodology to diagnose the important factors on the severe accident conditions

Yongju Cho, Sunhong Yoon
KEPCO E&C, Nuclear Technology Research Department, 269 Hyeoksin-ro, Gimcheon-si, 39660
rty1474@kepco-enc.com

1. Introduction

We have developed a diagnostic methodology using machine learning (ML) technology to figure out the important information including the break size, location, other remarkable events such as core uncover, relocation, reactor vessel failure and so on. When predicting the important facts, we found the fact that a more accurate prediction is possible if more parameters are taken into accounts. If all the plant's measurement parameters are used, the results of the prediction can be more accurate, but it seems not reasonable in ML because such an approaching method requires huge big data and very high capacity-level hardware system. So It is not economical. The measurement parameters in the power plant are expressed in our ML model as features, and we developed the diagnosis ML model to optimize the number of the features and applied it to predict the break sizes of LBLOCA accidents.

2. Methods and Results

2.1 Training data

The features to be used in machine learning were selected by considering the measurement parameters related to SAMG. So, we extracted 30 parameters (in ML, called features). The break sizes of the LBLOCA ranged from 6 inches to 16 inches with a spacing of 0.01 inch. Thereby, we had the dig data from the analyses of 1,400 cases, each feature (j) of each case was integrated over 60 seconds after scram using the following equation of (1) [1]. Then, we made Table 1 data set for ML in integrated 1400 cases and 30 parameters.

$$x_j = \int_{\Delta t}^{t_s+\Delta t} g_j(t)\, dt \qquad j = 1,2,3 \ldots\ldots 30 \qquad \textbf{(1)}$$

where $g_j(t)$ is a simulated signal, Δt is an integrating time span, and t_s is reactor scram time **[3]**

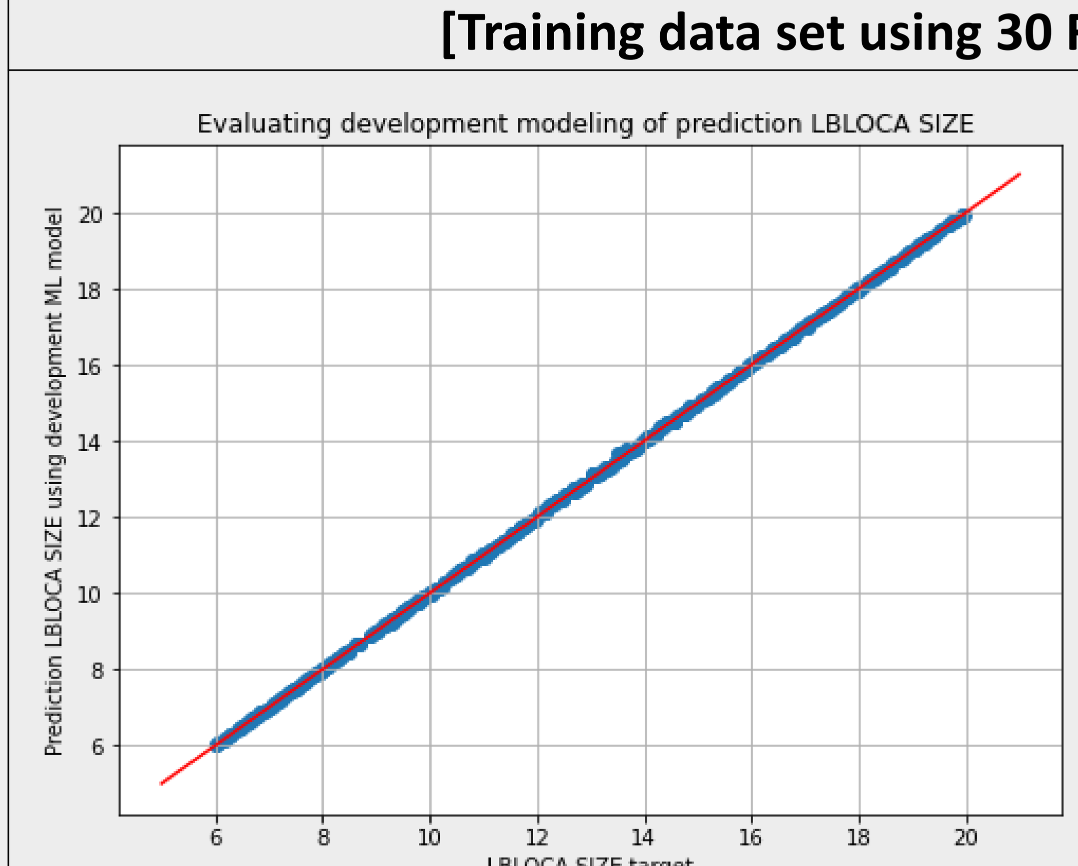
Table 1 : Data set for ML(1401rows × 32columns)

case	LOCA SIZE (Inch)	Pressure in pressurizer	Pressure in primary system	Temp in pressurizer compt	Temp in annular compt #4	Temp in annular compt #3
0	6.00	5.440798 e+08	5.440798e+ 08	20148.7761	20219.2849	20265.3181
1	6.01	5.439511 e+08	5.439511e+ 08	20150.2994	20220.8184	20267.0001
2	6.02	5.448275 e+08	5.448275e+ 08	20151.1197	20221.8602	20267.7797
1399	15.99	3.803337 e+08	3.803337e+ 08	23027.0653	23514.4052	23356.4910
1400	16.00	3.804734 e+08	3.804734e+ 08	23034.2771	23520.0511	23364.1599

2.2 Machine learning

Before training the ML models, preprocessing of null value is required. 70% of the existing data was used for training and the rest was used for evaluation. The evaluation criteria is MSE(Mean Squared Error), RMSE(Root Mean Squared Error), MAPE(Mean Absolute Percentage Error), R^2score. There are no absolute evaluation criteria for modeling evaluation indicators. In general, a good ML model has smaller MSE, RMSE, and MAPE values while larger R^2 scores. As a result, it was concluded that the model trained by 30 input features is reasonable model for LBLOCA size diagnosis (Table 2)

Table 2 : Performance development modeling when using 30 features [Training data set using 30 Features]



MSE	0.0008
RMSE	0.0284
MAPE	0.1745
R^2 score	0.9999

2.3 Important features extraction methodology

First, the MDI method was used to evaluate the previously presented development model. The next thing to do was making data set for each feature (Pressure in pressurizer, Pressure in primary system, Temp in pressurizer compt etc.…). Then, train the algorithm in the same way as before using the data set for each feature.

Table 3 showed the MDI rank of a model trained with 30 features as a data set, and MSE, RMSE, MAPE and R^2 score were modeling values trained with each feature as data set.

After 15 MDI rank features, one or more of the MSE, RMSE, MAPE, R^2 scores differed by at least several times. We found out that the features with low MDI rank are mostly poor MSE, RMSE, MAPE, R^2 scores. In summary, it was necessary to look at MDI MSE, RMSE, MAPE and R^2score. In this model We removed 14 features in order to optimize the number of the features. They were marked in red in the Table 3.

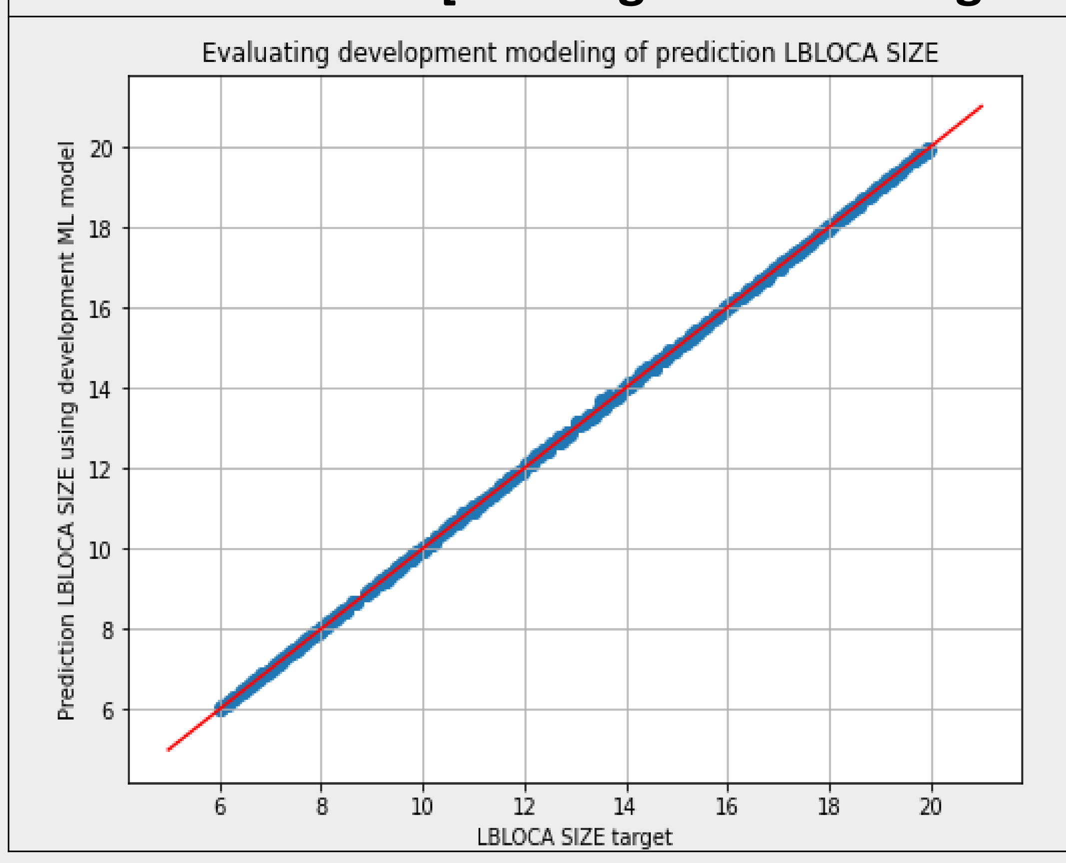
Table 3 : Evaluate each feature using developing modeling

[Training data set using each Feature]

MDI rank	Feature	MSE	RMSE	MAPE	R^2 score
1	Water level in Cavity	0.0274	0.1654	0.2852	0.9983
2	P in pressurizer	0.0689	0.2625	0.9957	0.9957
3	Core Exit T	1.9449	1.3946	8.0565	0.8797
4	Core collapsed water level	0.0093	0.0967	0.6781	0.9994
5	Pressure in CTMT dome	0.0016	0.0405	0.2338	0.9999
….	…..	…..	…..	…..	…..
15	T of gas in annular Compt SW-#1 El.100'	0.00264	0.0514	0.2585	0.9998
16	Water Temp in loop 4 cold leg	3.8151	1.9532	12.3814	0.7640
17	Water Temp in loop 2 cold leg	3.8030	1.9501	12.4084	0.7647
…..	….	….	….	….	….
20	Water Temp in loop 3 cold leg	2.1814	1.4770	8.4166	0.8650
21	Water Temp in loop 1 cold leg	2.0795	1.4426	8.4743	0.8713
22	Flow rate of ESF	10.4055	3.2258	25.8784	0.3562
…..	….	….	….	….	….

Table 4 is LBLOCA Size Prediction And Performance using machine learning through 16 features extracted by excluding those with lower MDI ranking or poor R^2 score, MSE, RMSE and MAPE values. After extracting them, the machine learning model was trained using this Training data. The results were similar to those of the model with 30 input features shown in Table 2.

Table 4 : LBLOCA Size Prediction And Performance using machine learning through optimizing the features [Training data set using 16 Features]



MSE	0.0010
RMSE	0.0329
MAPE	0.1910
R^2 score	0.9999

3. Conclusions

The purpose of this study was to optimize by reducing the number of the measurement parameters (or features) required for machine learning without compromising accuracy. We found that the random forest sampling method and the evaluation criteria such as MSE, RMSE, MAPE, R^2 score are useful to figure out the important ones out of all the considerable features. In the future, the optimized features will be derived by extending the prediction of another scenario shown in the progresses of severe accident.

REFERENCES

[1] Geon Pil Choi, Kwae Hwan Yoo, Estimate of LOCA Break Size Using Cascaded Fuzzy Neutral Networks, Nug. Eng. Technol. 49 (2017) 495-503