Development to Diagnose Model of Abnormal Status in Nuclear Power Plant Operation using Machine Learning Algorithms

Ho Sun Ryu[a]*, Kwang Nam Yu[b],Yun Goo Kim[a]
[a]*Korea Hydro and Nuclear Power Co., ltd, Central Research Institute, Daejeon, Korea*
[b]*Dacon Co., Seoul, Korea*
**Corresponding author : hosunryu@khnp.co.kr*

1. Introduction

When an abnormal status occurs in a nuclear power plant, the operator determines the abnormal condition and takes action according to each manual.

However, nuclear power plants have over 80 abnormal conditions. And there are more than 200 events inside the abnormal state and there are many driving variables. Therefore, the operator may have difficulty determining the abnormality. And, depending on the level of the operators, a difference may occur in judging and handling abnormal conditions.

So, KHNP CRI is conducting research projects to automatically determine abnormal conditions using artificial intelligence(AI).[1] In addition, in order to apply the latest artificial intelligence technology in the research project and secure the foundation for technological exchange with artificial intelligence experts, an artificial intelligence model development competition (hackathon) was held online.

In this paper, we introduce the hackathon competition promoted for the development of AI models and evaluate them on the developed AI models.

2. Generating train and test data

To develop an AI model, there must be a lot of data but it is difficult to develop an AI model only with power plant operation data because there are few cases where abnormal conditions occur in nuclear power plants.

So, the actual abnormal operation data and simulation data were used together. Using the full-scope simulator simulating the operation situation of a nuclear power plant, abnormal operation data was generated and used as AI training and testing data. The simulation data can be used as plant training data because the simulation values do not exactly match the actual values of the plant, but the trend is similar.[2]

Based on the Shin-Gori Unit 3, the abnormal status of the power plant were simulated in 21 out of 82. In addition, 21 abnormal states were separated into 198 cases, and 1610 data files were generated.

The format of the generated data is shown in Figure 1. There are a total of 5121 driving variables, and data for 10 minutes are acquired at 1 second intervals.

Time	Var1	Var2	...	Var5121	Label
1	0.529427	0.4363186	0.2115986	0.328125	0
2	0.529427	0.4363186	0.2115986	0.328125	0
3	0.529427	0.4356992	0.2115986	0.426822667	0
4	0.529427	0.4356992	0.2115986	0.426822667	0
5	0.529427	0.4356992	0.2115986	0.426822667	0
-	-	-	-	-	0
597	1.529427	0.4356992	0.2115986	0.426822667	3451
598	1.5289063	0.436009	0.2117513	0.407552	3451
599	1.5289063	0.436009	0.2117513	0.407552	3451
600	1.5289063	0.436009	0.2117513	0.407552	3451

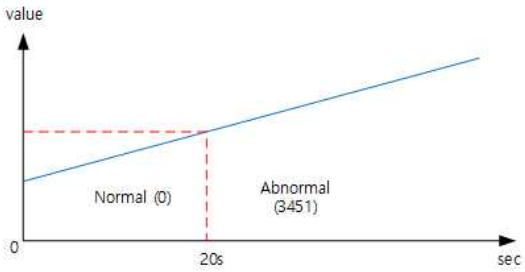

Fig. 1. Example of saved simulation data for abnormal state

In order to create a value similar to the actual power plant data, the plant noise pattern was extracted for each variable and applied to the simulation data.

For this, an exponential smoothing technique was applied. The standard deviation is obtained by the difference between the actual data and the flattening data, and noise is generated and applied to the simulator data to produce data as shown in Figure 2 below.

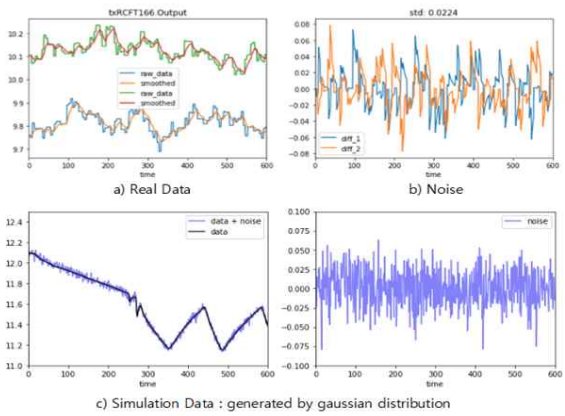

Fig. 2. Make simulation data induced noise

3. Development a Machine Learning Model

3.1 Operation of Hackathon

Hackathon is a compound word of hacking and marathon, which means an event or data competition that focuses on short-term intensive work in the software field. Even in overseas cases, artificial intelligence has developed and applied many excellent algorithms through data competitions.

The hackathon competition was held online for a month. For the operation of the competition, a discussion page, a code sharing page, a submission page, and a leaderboard page were operated as shown in Figure 3. The leaderboard automatically evaluates the data collected by the participants and displays the ranking in real time.[3]

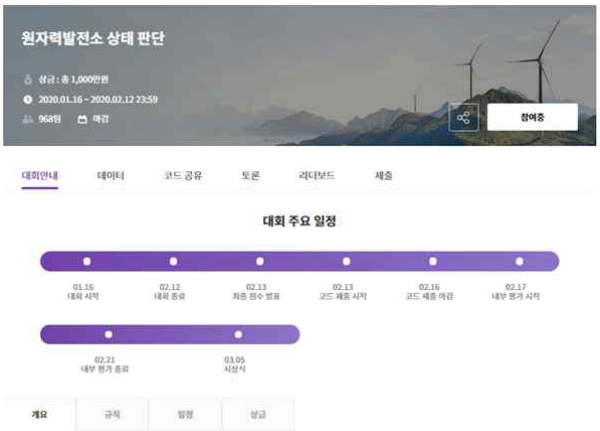

Fig. 3. Online Site of Hackathon

Determining the abnormal state is a classification problem and it can generate a AI model using a decision tree algorithm during machine learning. So, we used the RandomForest Algorithm of the decision tree to create and distribute reference code to encourage participation and lower the barrier to entry into the competition. And then, a video explaining the base line code was made and published on the YouTube channel.[4]

3.2 Evaluation Model

The following LogLoss score was used for ranking.

$$LogLoss = -\frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M} y_{n,m}\log(p_{n,m})$$

$y_{n,m}$: Answer probability
$p_{n,m}$: Prediction value
M : Total number of labels
N : Total number of test data

Accuracy is to determine whether the correct answer is correct. If the final value is the same regardless of the probability of the correct answer, the same evaluation result is produced. On the other hand, LogLoss reflects the probability of getting the correct answer, so you can evaluate the model more precisely. We evaluate based on the converted value using the Log function to utilize the probability value.

The reason for converting the probability value to the Log function is to give more penalty as the probability is predicted lower. If the correct answer is completely wrong, a high penalty is given, and if the correct answer is set to 100% probability, the penalty is not imposed.

After the competition period was over, the program source was provided to the top 8 teams in real-time ranking, the algorithms used were reviewed, and the performance test was conducted on additional data that was not disclosed, and the final 5 teams were selected.

The final score adds up 50% of the published data and 50% of the private data. In order to evaluate model performance for short data, the total number of non-public data is 154, and the same data is used for 60 seconds, 50 seconds, 40 seconds, and 30 seconds, and the power plant data is weighted to give a higher score when correct answers are given.

3.3 Result of Hackathon

The total number of participants was 963, with the largest number of data competitions held in Korea. The top 5 teams' LogLoss scores and accuracy are shown in Figure 4. In addition, Table 1 summarizes the applied models.

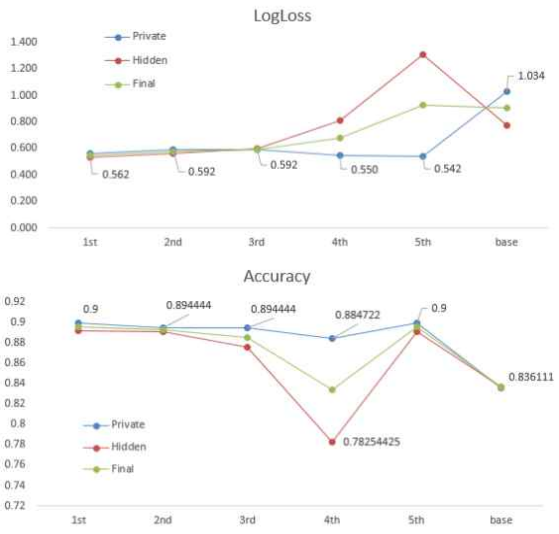

Fig. 4. Result of AI Model

The first team predicted the simulated abnormal status with 90% accuracy, and showed a high accuracy of more than 90% without compromising performance even in a private data set not disclosed during the competition.

Table 1. Result of AI Model

	1st	2nd	3rd	4th	5th	Base Line
Model	LightGBM(12)	LightGBM(1)	LightGBM(10) + XGBoost(4)	LightGBM(15)	LightGBM(4) + XGBoost(7)	RandomForest(1)
Validation	4-Fold CV	1000 epoch training	10-Fold CV 4-Fold CV	Reader Board	-	-
Prediction	Soft voting	MinMax correction	LightGBM, XGBoost 7:3 ensenble	Average value	Average value	-
LogLoss	0.568339	0.576753	0.594981	0.704011	0.925579	0.906672
Accuracy	0.9	0.892488	0.885092	0.833633	0.895266	0.836695

And all 5 teams used the latest AI model LightGBM to limit the high-performance AI model. LightGBM is a machine-learning algorithm developed by Micro Software, and it has recently attracted much attention in terms of algorithm speed and accuracy even overseas. [5] In addition, various machine learning techniques such as XGBoost, K-fold cross-validation, and ensemble were used.

4. Conclusion

In the future, we will additionally proceed with simulation using the full-scope simulator and the abnormal status data of the actual power plant. The AI model developed through hackathon will be corrected and supplemented using additional generated data.

Developing an AI model requires a lot of resources, such as data, time, and computer resources. We developed a high-performance AI model in a short time through the hackathon competition. This is expected to be a good precedent business model that applies AI models in various fields as well as power plants in the future.

REFERENCES

[1]Y.G. Kim, et al, "Consideration on data mapping of convolutional neural networks to diagnose abnormal status in nuclear power plant operations," KNS spring meeting 2018

[2]Y.G. Kim, "Development of Convolutional Neural Networks to Diagnose Abnormal Status in Nuclear Power Plant Operation" , KNS spring meeting, 2019

[3]https://www.dacon.io/competitons/official/23551/overview

[4]https://www.youtube.com/watch?v=TyO9yQubqkg&feature =youtu.be

[5]G.Ke, T.Finley, "LightGBM : A Highly Efficient Gradient Boosting Decision Tree", NIPS 2017

ACKNOWLEDGEMENT

This work was supported by the Energy Efficiency & Resources of the Korea Institute of Energy Technology Evaluation and Planning(KETEP) grant funded by the Ministry of Trade, Industry and Energy. (No. 20171510102040)