Detection of Untrained Class for Accident Diagnosis Model with Open Set Recognition Method

Seung Geun Kim [a*], Young Ho Chae [b], Seo Ryong Koo [b]
[a]*Applied Artificial Intelligence Section, Korea Atomic Energy Research Institute, Daedeok-daero 989beon-gil, Yuseong-gu, Daejeon 34057, Republic of Korea*
[b]*Advanced Instrumentation and Control Research Division, Korea Atomic Energy Research Institute, Daedeok-daero 989beon-gil, Yuseong-gu, Daejeon 34057, Republic of Korea*
[*]*Corresponding author: sgkim92@kaeri.re.kr*

Keywords : event diagnosis, accident diagnosis, untrained class detection, open set recognition, nuclear power plant

1. Introduction

To ensure the safety and efficiency of nuclear power plants(NPPs), operators monitor various instrumentation signals and alarms, and conduct diagnostic actions if any event or accident occurs. Since accurate diagnosis is essential for establishing proper mitigation strategy, and inappropriate diagnosis under harsh condition may result in severe consequences, event/accident diagnosis is regarded as one of the important tasks of operators in NPPs.

Although procedures for event/accident diagnosis are well-defined, still there exists a possibility of human error occurrence due to various factors including but not limited to time pressure for diagnosis and excessive information inflow. In this regard, there have been efforts for the development of artificial intelligence(AI)-based event/accident diagnosis models. Especially, due to the rapid advancement of AI technology based on deep neural network(DNN), recently developed event/accident diagnosis models are showing outstanding performances[1, 2].

Most of previously developed AI-based classification models, including NPP event/accident diagnosis models have common drawback that they always deduce output among the pre-defined classes, although there exists a possibility that the input is irrelevant to any of them. However, due to various reasons such as limitation of simulators or limitation on the amount of data, most of diagnosis models are unable to cover entire possible event/accident scenarios. Under the unexpected or untrained situations, the model's confident-yet-inaccurate output could make operators to be confused, leading to the inappropriate responses.

Therefore, for the practical application of AI-based event/accident diagnosis model, the concept of open set recognition should be considered that grants the ability for detecting untrained classes to the model. In this study, the applicability of OpenMax[3]-which is one of the representative open set recognition method-for the NPP diagnosis model is investigated. For the experiments, simple accident diagnosis model is developed based on data acquired from compact nuclear simulator(CNS)[4], and OpenMax method is applied for the detection of untrained class(i.e. intentionally neglected class of data during training).

The rest of paper is organized as follows. In chapter 2, brief explanation about OpenMax method is provided. In chapter 3, processes of experiments and the corresponding results are described. Chapter 4 summarizes and concludes the paper.

2. Method: OpenMax

For the open set recognition in NPP accident diagnosis model, OpenMax[3] method is applied in this study. OpenMax is one of the discriminative open set recognition methods, which focus on defining precise decision boundaries of trained classes and then setting thresholds to detect untrained classes. The method is simple to implement since the underlying concepts are intuitive, and easy to apply as the method does not involve model configuration changes.

There are two steps need for the application of OpenMax, that are preparation step and execution step. Preparation step is for establishing standards for detecting untrained classes based on training data, while execution step is for conducting actual untrained class detection for given input. During these steps, the method utilizes activation vector(AV), which is a set of node values in output layer before activation.

There are three sub-steps in preparation step, which are P1) data sorting, P2) AV profiling, and P3) extreme value fitting.

P1) data sorting: in this step, only correctly classified training data by the model are sorted and separated according to the class.

P2) AV profiling: in this step, mean AV and mean distance between mean AV and AVs of each data are calculated for each trained class. The distance is calculated with Euclidean-cosine distance measure, which can be represented as follows.

$$EC\_dist = Euclidean\_dist \times (1 - cos\_sim)$$
$$Euclidean\_dist = \|(V_1 - V_2)\|_2$$
$$Cos_{sim} = \|V_1\|_2 \cdot \|V_2\|_2$$

Where EC_dist is Euclidean-cosine distance, $Euclidean\_dist$ is Euclidean distance, Cos_sim is cosine similarity, V_1 and V_2 are given AVs and $\|\cdot\|_2$ implies the L2-norm.

P3) extreme value fitting: in this step, distribution of calculated distances is estimated based on extreme value theory(EVT) for each class. In general, Weibull distribution is applied for fitting. For the distribution estimation, top η (hyperparameter) farthest AVs from mean AV of each class are selected as extreme values. Probability density function(PDF) of Weibull distribution can be represented as follows.

$$f(x;\lambda,k)=\begin{cases}\frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}e^{-(x/\lambda)^k}, & x\geq 0\\ 0, & x<0\end{cases}$$

Where k and λ are positive shape and scale parameter of the distribution, respectively.

There are also three sub-steps in execution step, which are E1) AV calculation, E2) probability calculation, and E3) AV revision.

E1) AV calculation: in this step, AV of the given input and its distance from mean AV of each trained class are calculated.

E2) probability calculation: in this step, based on the estimated distance distribution at step P2), probability that the distance to be same or lower than the calculated distance from mean AV is calculated for each trained class. If the AV of input is similar to the specific class' mean AV, the calculated probability for corresponding class would be low. If the AV of input is highly different from specific class's mean AV, the calculated probability for that class would be high. The calculated probability for class k is denoted as ω_k, and utilized in the next sub-step.

E3) AV revision: in this step, based on the calculated probability at step E2), AV of the input is revised, and new element for untrained class is added. For class k, revision of AV is conducted as follows.

$$AV_k'=(1-\omega_k)\times AV_k$$
$$AV_0'=\sum_{i\in K}(\omega_i\times AV_i)$$
$$K=1,2,\dots N\ (\text{N: number of trained classes})$$

Where AV_k and AV_k' represents the element correspond to the trained class k before and after the revision, respectively. AV_0' represents the added element correspond to the untrained class after the revision.

Revised classification probabilities can be easily calculated based on revised AV and Softmax function. Revised classification probability for class k and untrained class can be represented as follows.

$$Pr(k)=\frac{\exp(AV_i')}{\exp(AV_0')+\sum_{i\in K}\exp(AV_i')}$$
$$Pr(untrained)=\frac{\exp(AV_0')}{\exp(AV_0')+\sum_{i\in K}\exp(AV_i')}$$
$$K=1,2,\dots N\ (\text{N: number of trained classes})$$

Where $Pr(k)$ is the revised classification probability for class k, and $Pr(untrained)$ is the classification probability for untrained class.

3. Experiments

To conduct experiments, a simple NPP accident diagnosis model is developed in this study. Experiments are conducted through the following steps: data acquisition and preprocessing, model development and training, and application of OpenMax method.

3.1 Data acquisition and preprocessing

For the development of accident diagnosis model, data is acquired from simulations with compact nuclear simulator(CNS)[4]. Reference plant of CNS is Westinghouse 3-loop MWe pressurized water reactor(PWR). Loss of coolant accident(LOCA) from loop 1/2/3 cold/hot leg(break sizes from 15 to 35cm^2, with 5cm^2 interval), steam generator tube rupture(SGTR) from loop 1/2/3(break sizes from 4 to 20cm^2, with 4cm^2 interval), and main steam line break(MSLB) from loop 1/2/3 inside/outside containment(break sizes from 500 to 1000cm^2, with 100cm^2 interval) accident scenarios are considered. Simulations are conducted for 20 minutes from reactor trip and signals from 5 to 15 minutes are used as data. As variables, totally 19 kinds of variables are acquired during simulation(refer Table I).

After the simulation, minimum-maximum scaling(min-max scaling) is applied to set the range of variables between 0 and 1, and unit data with 5 minutes length is generated with 10 seconds interval.

Table I: list of acquired variables during simulation

Acquired variables	Units
Cold leg temperature (loop 1/2/3)	℃
PZR pressure	kg/cm^2
PZR level	%
S/G pressure (loop 1/2/3)	kg/cm^2
S/G level – Wide range (loop 1/2/3)	%
Feedwater flow (line 1/2/3)	ton/hr
Steamline flow (line 1/2/3)	ton/hr
Containment radiation	mRem/hr
Secondary cooling system radiation	μCi/cc

Total number of generated unit data is 558 for LOCA, 279 for SGTR, and 744 for MSLB accident scenario. Among them, 70% are used for training, 15% for validation, and rest 15% for testing.

3.2 Model development and training

Based on the acquired data, DNN-based simple accident diagnosis model that consist of five fully-connected feed-forward layers is developed and trained. Except output layer with Softmax activation function, ELU(exponential linear unit) activation function is applied to all other layers. Fig. 1 is a schematic of developed accident diagnosis model.

Models are separately developed with changing the neglecting accident class. First model is trained with neglecting MSLB data(referred as case 1), while second model is trained with neglecting SGTR data(referred as case 2) and third model is trained with neglecting LOCA data(referred as case 3).

As a result, 3 kinds of models are developed with changing neglected accident class. All developed models have achieved 100% accuracy for the training/validation/testing data.

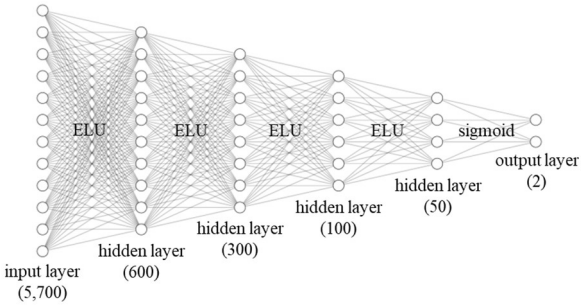

Fig. 1. Schematic of developed accident diagnosis model

3.3 Application of OpenMax method

For the developed accident diagnosis models, OpenMax methods are applied to detect untrained class. Experiments are conducted with changing η value from 10 to 300.

Table II represents the results that achieved highest mean classification accuracy of trained and untrained classes. Negative values in brackets represent the decrement of classification accuracies due to the application of OpenMax method. Fig. 2, 3 and 4 represent the changes of classification accuracies according to η value for each case.

Table II: Classification accuracies

	LOCA	SGTR	MSLB
Case 1 U*: MSLB	98.92% (-1.08%)	100.00% (-0%)	99.06% (Untrained)
Case 2 U: SGTR	100.00% (-0%)	100.00% (Untrained)	100.00% (-0%)
Case 3 U: LOCA	100.00% (Untrained)	99.28% (-0.72%)	98.66% (-1.34%)

*'U' represents untrained class

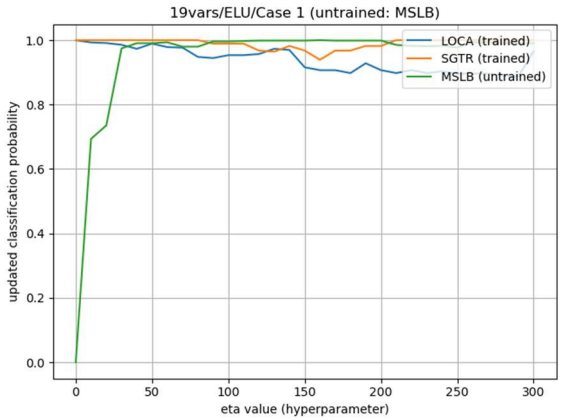

Fig. 2. Classification accuracies for case 1 with changing η value (best when η = *50*)

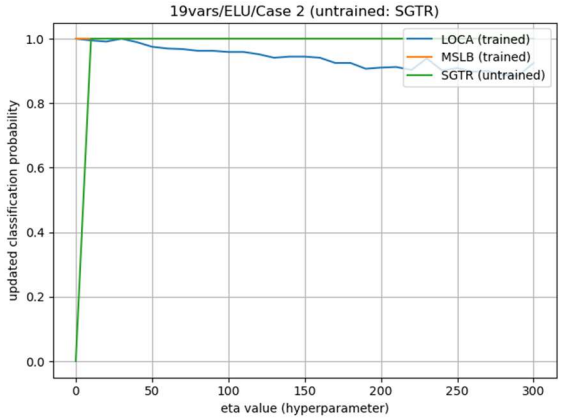

Fig. 3. Classification accuracies for case 2 with changing η value (best when η = *30*)

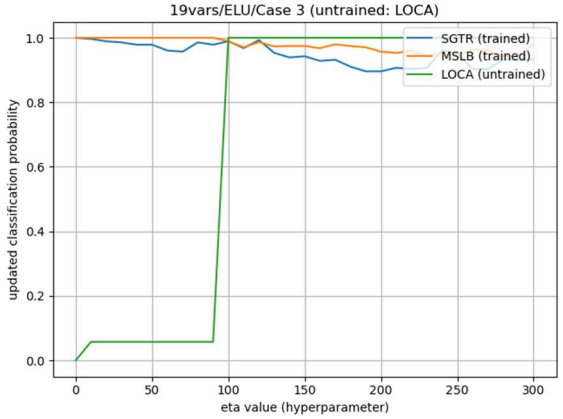

Fig. 4. Classification accuracies for case 3 with changing η value (best when η =*120*)

From the experiments, it is revealed that the OpenMax method is able to detect untrained class with over 99% accuracy in every cases, with about 1% classification accuracy drop for trained classes.

For the cases with changing untrained classes, η values correspond to the best performances are different. In general, increasing the η value tends to result in increased probability for untrained class detection, while it also decreases the classification performance of trained classes. Setting of proper η value is important for

untrained class detection, since the performances for untrained class detection and trained class classification are in trade-off relations in macroscopic view.

In practice, experiments for 'actual' untrained class detection would be difficult since there would be no available data for untrained class. Alternatively, sub-optimal η value can be found based on the experiments with intentional neglection of specific class within acquired data, similar to the experiments conducted in this study,

4. Conclusions

In this study, OpenMax method, which is one of the open set recognition method is applied for the detection of untrained class in NPP accident diagnosis model. For the experiments, DNN-based simple accident diagnosis model is developed based on data acquired from CNS, and OpenMax method is applied to check the performance for untrained class detection and classification performance drop for trained class during the application. The experiment results have shown that the OpenMax method is capable of untrained class detection with high accuracy, with acceptable classification performance drop for trained classes.

As future works, the effects of model structure and input space complexity to the untrained class detection will be investigated. In addition, comparison between various open set recognition methods will be conducted.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (No. RS-2022-00144150)

REFERENCES

[1] Y. H. Chae, et al., "Graph Neural Network based Multiple Accident Diagnosis in Nuclear Power Plants: Data Optimization to Represent the System Configuration", Nuclear Engineering and Technology, Vol. 54, No. 8, pp. 2859-2870, 2022.

[2] J. H. Shin, et al, "Approach to Diagnosing Multiple Abnormal Events with Single-event Training Data", Nuclear Engineering and Technology, 2023 (available online),m ISSN 1738-5733, DOI:10.1016/j.net.2023.10.033

[3] A. Bendale, and T. E. Boult, Towards Open Set Deep Networks, Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR2016), pp. 1563-1572, 2016.

[4] KAERI, "Advanced compact nuclear simulator textbook", KAERI, Korea Atomic Energy Research Institute(KAERI) Nuclear Training Center, Daejeon, South Korea, Technical Report., 1990.