Robust Nuclear Power Plant AI Modeling: A Cross-Type Simulation using Knowledge Distillation

Young Ho Chae[a], Seung Geun Kim[a], Seo Ryoung Koo[a*]
[a] *Korea Atomic Energy Research Institute, 111 Daedeok-daero 989 beon-gil, Daejeon, 34057*
[*]*Corresponding author: srkoo@kaeri.re.kr*

***Keywords :** Robustness, Cross-Type Simulation, Knowledge distillation

1. Introduction

Deep learning methodologies, which utilize artificial neural networks, are becoming increasingly prevalent in nuclear energy research. This prevalence is often attributed to their ability to outperform traditional machine learning techniques when ample data is available. Within the field of Probabilistic Safety Assessment (PSA), deep learning has found applications such as scenario optimization, as evidenced by the work of Bae et al.[1] Meanwhile, in the simulation domain, physics-informed neural networks have been employed for surrogate modeling, a technique highlighted in the research conducted by Antonello et al.[2]

As can be seen from recent research, neural network-based methodologies have been successfully employed across various fields. However, they present two main challenges. The first challenge is the size and complexity of the model, which necessitates significant computational resources. This is not only true for the training phase but also for deploying the trained neural network. The second challenge lies in data dependency. In several accident diagnosis studies, such as those by Chae et al.[3] and Kim et al.[4], real-world testing in nuclear power plants under abnormal or emergency conditions is infeasible. Consequently, these studies have had to rely on simulated data. If the robustness of the artificial neural network is not adequately ensured, difficulties arise in applying it to real-world scenarios, limiting its applicability to simulated environments.

To address these issues, we propose a method to both reduce the model's size and enhance its robustness by employing a distillation methodology. This approach aims to provide a more practical and efficient solution for leveraging neural networks in complex environments.

2. Knowledge Distillation

Knowledge distillation is a process where a lightweight student model is trained using the learning process and outcomes of a more complex teacher model (Hinton et al. [5]). The teacher model is typically a sophisticated and highly accurate neural network, and the essence of knowledge distillation lies in guiding the student model's learning with newly created information from the teacher model.

Knowledge distillation consists of two types: Feature-based knowledge distillation, which distills the components within an artificial neural network according to the type of data being distilled, and Response-based knowledge distillation, which utilizes the output results of the artificial neural network. Together, these two methods are referred to as relation-based distillation.

To illustrate the process of response-based distillation, consider designing a neural network to classify dogs, cats, and birds. The training data might be organized in Table 1, and a well-trained network would utilize a softmax classifier (Eq. 1) to obtain the results shown in Table 2.

Notations

Symbol	Definition
i	Class
q_i	Expected probability
τ	Temperature
x	Input
y	Output
D	Domain, Codomain
L_{KD}	Kullback-leibler loss
S	Student model
T	Teacher model
θ	Trainable parameters (e.g. weight)
λ	Constant
L_{CE}	Cross-entropy loss

Table 1: Example dataset

	Label		
Input Image	Dog	Cat	Bird
Dog	1	0	0
Cat	0	1	0
Bird	0	0	1

$$q_i = \frac{\exp(z_i)}{\sum_j \exp(z_j)}$$

Eq. 1

Table 2: Example classification results

Label (Test Data)	Probability		
	Dog	Cat	Bird
Dog	0.9	0.07	0.03
Cat	0.03	0.95	0.02
Bird	0.006	0.004	0.99

The key insight here is the inference result of the Teacher model. For instance, when test data with the 'Dog' label is processed, the model outputs 'Dog' but also generates information indicating that the image is more similar to a cat than a bird. Similarly, with 'Cat' labeled data, information could be gleaned that the image is closer to a dog than a bird. In the Student model, both the original hard target (e.g., [1, 0, 0]) and additional soft target information created by the Teacher model (e.g., [0.9, 0.07, 0.03]) can be used. In practice, Temperature concept is utilized to smooth the output (Eq. 2).

$$q_i = \frac{\exp\left(\frac{z_i}{\tau}\right)}{\Sigma_j \exp\left(\frac{z_j}{\tau}\right)} \qquad \text{Eq. 2}$$

This enables the creation of a more compact and robust model. The process can be further described by the loss function in Eq. 3.

$$\begin{aligned} &L \\ &= \Sigma_{x,y\in D} L_{KD}\big(S(x,\theta_s,\tau),T(x,\theta_T,\tau)\big) \\ &+ \lambda L_{CE}(\hat{y}_s, y) \end{aligned} \qquad \text{Eq. 3}$$

The benefits of utilizing knowledge distillation (response-based distillation) are as follows in terms of robustness.

1. Mitigation of overfitting: In the case of general neural networks, the correctness of the answer is important, but since Student network mainly evaluates the similarity with the distilled information, the overfitting problem is reduced.

2. Denoising features: the student might learn to focus on the most relevant and robust features of the data, effectively denoising or discarding less pertinent information.

3. Condensation of complexity: Response-based distillation can allow the student to capture the essence of these patterns which shown in the teacher network in a more compact form, potentially leading to improved robustness.

3. Robust NPP AI Modeling (Experimental Setup)

Three distinct simulators—namely the Compact Nuclear Simulator (CNS), PCTRAN Simulator, and 3-Key Master Simulator—were employed for data acquisition. This was done to assess the effectiveness of making the model more compact and robust through the Knowledge Distillation (KD) technique.

- Compact Nuclear Simulator (CNS): A full-scope simulator developed for educational, training, and accident response purposes. It simulates various accident situations within a Westing House 3-Loop Pressurized Water Reactor, designed by Westinghouse, with three circulation loops.

- PCTRAN Simulator: Also a full-scope simulator, PCTRAN is designed to perform transient analysis by simulating various accident situations. It simulates a power plant with an electrical output of 1400 MWe and two circulation loops.

- 3-Key Master Simulator: Like CNS, the 3-Key Master Simulator is a full-scope simulator developed for education and training of power plant operators. It also assists in establishing accident response strategies by simulating various accident situations. This simulator also replicates a power plant with an electrical output of 1400 MWe and two circulating loops.

Simulation results were gathered for three kinds of accident situations: Loss of Coolant Accident, Steam Generator Tube Rupture, and Main Steam Line Break, using the different simulators. From the collected data, the experiment was designed as illustrated in Fig.1.

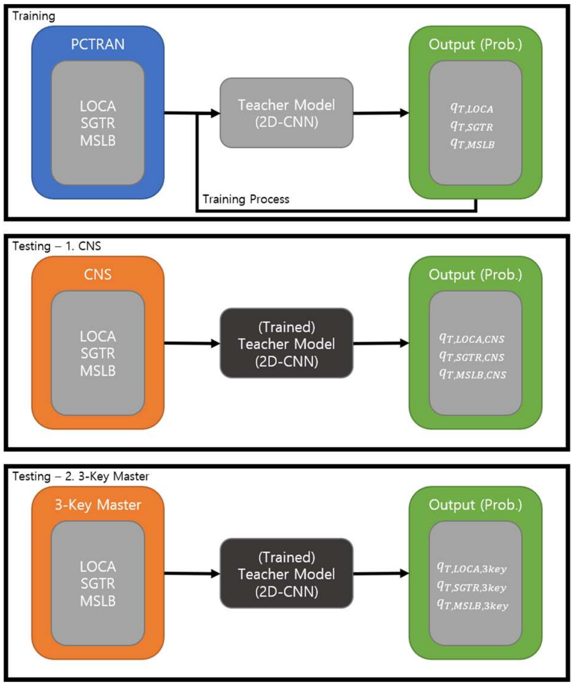

Fig. 1 Diagnosis w/o KD

Initially, the Teacher model was trained to diagnose the three aforementioned accidents using data obtained from the PCTRAN Simulator. Once sufficiently trained, the output probability of the Teacher model was calculated for both CNS ($q_{T,LOCA,CNS}$, $q_{T,SGTR,CNS}$, $q_{T,MSLB,CNS}$) and 3-Key Master Simulator data ($q_{T,LOCA,3Key}$, $q_{T,SGTR,3Key}$, $q_{T,MSLB,3Key}$).

The experimental design employing the KD methodology is depicted in Fig. 2.

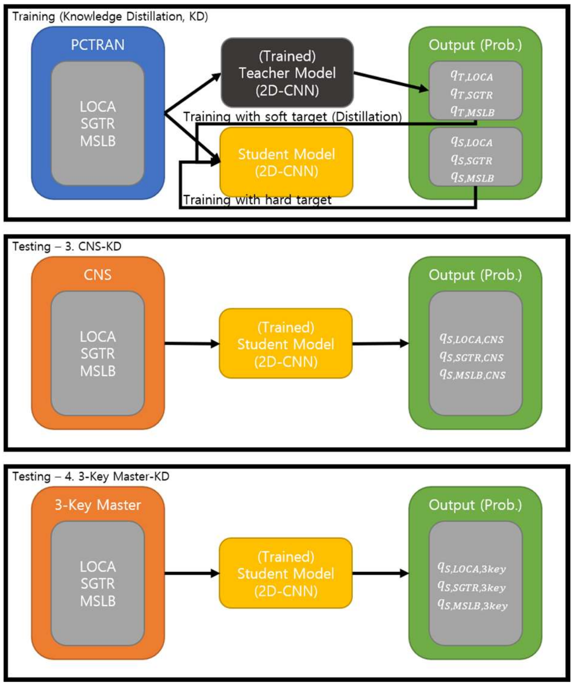

Fig. 2 Diagnosis w/ KD

In the experiment, the output probability result of the trained Teacher model serves as a soft target to further train the Student model. Once the Student model is adequately trained, the output probability for the CNS ($q_{S,LOCA,CNS}$, $q_{S,SGTR,CNS}$, $q_{S,MSLB,CNS}$) and 3-Key Master Simulator data ($q_{S,LOCA,3key}$, $q_{S,SGTR,3key}$, $q_{S,MSLB,3key}$) is calculated in a manner similar to that done for the Teacher model.

4. Conclusion and Future Works

In the context of both the CNS and 3-Key Master Simulators, parameters such as $q_{T,LOCA,3Key}$, $q_{T,SGTR,3Key}$, $q_{T,MSLB,3Key}$ and $q_{S,LOCA,3key}$, $q_{S,SGTR,3key}$, $q_{S,MSLB,3key}$, as well as diagnosis accuracy of student and teacher model, can be instrumental in demonstrating the model's robustness to various types of data. Given that PCTRAN and the 3-Key Master Simulators have similar model structures, we anticipate that $q_{T,LOCA,3Key}$, $q_{T,SGTR,3Key}$, $q_{T,MSLB,3Key}$ and $q_{S,LOCA,3key}$, $q_{S,SGTR,3key}$, $q_{S,MSLB,3key}$, can be employed to assess robustness with respect to analogous model types. Conversely, when considering PCTRAN and CNS, since the cores are distinct, we expect that $q_{T,LOCA,CNS}$, $q_{T,SGTR,CNS}$, $q_{T,MSLB,CNS}$ and $q_{S,LOCA,CNS}$ $q_{S,SGTR,CNS}$, $q_{S,MSLB,CNS}$ could be utilized to gauge the robustness against different type simulators.

In the future, we plan to evaluate the feasibility of utilizing the knowledge distillation structure in the nuclear field by evaluating the performance for more diverse reactor types.

Acknowledgement

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] Bae, J., Park, J.W. and Lee, S.J. (2022) 'Limit surface/states searching algorithm with a deep neural network and Monte Carlo dropout for nuclear power plant safety assessment', Applied Soft Computing, 124, p. 109007. Available at: https://doi.org/10.1016/j.asoc.2022.109007.

[2] Antonello, F., Buongiorno, J. and Zio, E. (2023) 'Physics informed neural networks for surrogate modeling of accidental scenarios in nuclear power plants', Nuclear Engineering and Technology, 55(9), pp. 3409–3416. Available at: https://doi.org/10.1016/j.net.2023.06.027.

[3] Chae, Y.H. et al. (2022) 'Graph neural network based multiple accident diagnosis in nuclear power plants: Data optimization to represent the system configuration', Nuclear Engineering and Technology, 54(8), pp. 2859–2870. Available at: https://doi.org/10.1016/j.net.2022.02.024.

[4] Kim, J.M. et al. (2020) 'Abnormality diagnosis model for nuclear power plants using two-stage gated recurrent units', Nuclear Engineering and Technology, 52(9), pp. 2009–2016. Available at: https://doi.org/10.1016/j.net.2020.02.002.

[5] Hinton, G., Vinyals, O. and Dean, J. (2015) 'Distilling the Knowledge in a Neural Network'. arXiv. Available at: http://arxiv.org/abs/1503.02531.