Transactions of the Korean Nuclear Society Spring Meeting
Jeju, Korea, May 18-19, 2023

Signal validation in nuclear power plant accidents using a diffusion model

Jeonghun Choi*, Seung Jun Lee
Ulsan National Institute of Science and Technology, 50 UNIST-gil, Ulju-gun, Ulsan 44919, Korea;
**Corresponding author: jhchoi@unist.ac.kr*

1. Introduction

Field sensors are critical for maintaining the stability of nuclear power plants (NPPs), with deviations in these measurements having the potential to interrupt all operational processes, including safety feature actuation and operator awareness monitoring. In situations where immediate reactions are required, incorrect signals can have a significant impact on NPP safety, especially in NPP accident conditions. Faulty signals have contributed to well-known NPP accidents, such as the Three-Mile Island and Fukushima accidents. Signal validation is crucial for NPP safety and reliability.

Sensor errors can occur due to various factors such as sensors/transmitters, transducers, device software, and hardware defects. Sensor deviations have diverse types depending on their causes, with some drastic changes that induce unnecessary safety feature actuation, while others gradually deviate signal values. Design-basis accidents (DBAs) are considered in NPP system designs and are classified according to the location of the malfunctioning system, exhibiting different trends depending on the type of accident.

Advanced techniques are necessary to handle complex signal patterns in NPP accidents. Generative models, a novel artificial intelligence technique, can be applied to signal validation in accidents, with a signal fault detection system developed in this research using generative models for signal reconstruction and analysis. The development of signal reconstruction methods with generative models capable of learning the distribution of various accident data will be evaluated in terms of fault detection accuracy. This research aims to provide a comprehensive approach to address the challenges of sensor signal validation in NPP accidents, ensuring accurate diagnosis by plant operators and diagnostic tools.

2. Neural net-based signal validation method

This manuscript discusses signal validation methods using neural networks, which can be supervised or unsupervised [1]. The former has been applied to sensor fault detection in nuclear power plants using long short-term memory networks with good results but suffers from high learning costs and inflexibility. The latter employs an autoregressive process to reconstruct input signals and can be more flexible in its application. Two sequences consist of unsupervised learning-based signal validation: signal reconstruction and residual analysis [2]. In signal reconstruction, a model learns signal relationships by weight updates and generates a reconstructed signal from measured signal values. If there is a faulty signal, the fault detection model reconstructs the signal with diminishing fault influence by trained weights. Finally, sensor states are estimated by comparing the reconstructed and original signals.

2.1 Unsupervised learning with latent extraction

Unsupervised learning is a machine learning approach that discovers patterns in unlabelled data without guidance. It enables machines to learn similarly to humans, constructing a model of the environment and producing desirable content. An autoencoder, a common unsupervised learning-based model, learns to encode data by compressing it to latent representation and then decoding it back to its original form. It consists of an encoder and a decoder that compress and reconstruct the input data, respectively. The goal of an autoencoder is to capture essential features of the input data to reconstruct it with minimal loss of information.

In the case of multivariate time series data, an autoencoder can be used as an auto-regressive model to predict future values. The autoencoder is trained on a set of time series data to capture the temporal dependencies between variables, allowing it to predict future values. Variational autoencoders (VAEs) are generative models that learn a probability distribution of encoded representations to generate new data similar to the training data [3]. There are various types of autoencoders, each with unique architectures and objectives.

2.2 Denoising diffusion probabilistic model (DDPM)

Generative models have been extensively studied in recent years, with success in image generation tasks. This research has expanded to audio, text-to-speech, and forecasting tasks. The Denoising Diffusion Probabilistic Model (DDPM) or diffusion model is a latent model similar to VAE, which exhibits the highest performance among data generative models [4,5]. DDPM follows two processes.

The diffusion process is an iterative process of injecting interred noise into the original data leading to totally different (Independent) noise data. The basic principles of the diffusion process follow below eq.(1) and (2).

$$q(x_{1:T}^n|x_0^n) = \prod_{t=1}^{T} q(x_t^n|x_{t-1}^n) \; , \qquad (1)$$

$$q(x_t^n|x_{t-1}^n) = \mathcal{N}(\sqrt{1-\beta_t}x_{t-1}^n, \beta_t I) \; , \qquad (2)$$

intended noise injection follows the determined noise schedule including noise scale and diffusion rate, and variance of normal distribution.

The denoising process is a process of restoring the original data through the reverse process of starting from the noise data generated in the diffusion process following eq. (3) and (4).

$$p_\theta(x_{0:T}^n) = p(x_T^n)\prod_{t=1}^{T} p_\theta(x_{t-1}^n|x_t^n), \quad (3)$$

$$p_\theta(x_{t-1}^n|x_t^n) = \mathcal{N}(x_{t-1}^n; \mu_\theta(x_t^n, t)\Sigma_\theta(x_t^n, t)), \quad (4)$$

The diffusion model is an autoregressive model that transforms a noise vector to generate samples. Its loss function comprises reconstruction, regularization, and diffusion losses, with the first two being similar to VAE. The diffusion loss term drives the data generation process based on eq. (5).

$$Loss_{Diffusion} = D_{KL}(q(z|x_0)\|p_\theta(z) - E_q[logP_\theta(x_0|x_1)] + \sum_{t=2} D_{KL}(q(x_{t-1}|x_t, x_0)\|p_\theta(x_{t-1}|x_t)), \quad (5)$$

The diffusion model can generate contextual data by calculating multiple diffusion steps with more latent variables than VAE. We developed a signal validation system using this model and compared its performance in detecting sensor faults to prior techniques. The diffusion model was trained and tested using normal and sensor fault data from accident simulations.

3. Diffusion model implementation

3.1 Diffusion process on a multivariate signal

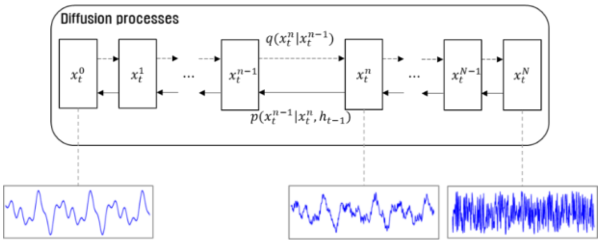

Fig. 1 Diffusion/denoising process in the diffusion model.

We developed a diffusion process (Figure 1) for NPP accident time series data. Using sensor signals in the model transforms them into different noises. An artificial neural network can be trained to invert this process and restore the original signal.

3.2 Diffusion model for signal validation

Based on the diffusion process, we developed a recurrent neural network (RNN) with a gated recurrent unit based on signal diffusion structure, which improved the performance of the RNN. The RNN was trained on hidden weights and underwent denoising steps of the diffusion model. The signal reconstruction system was built using RNN and reconstruction network structures. The RNN trained signal reconstruction information through the diffusion model, while the reconstruction network used a 1D-CNN-based regression model to develop signals. The training sequence of the signal validation system started by training the RNN with a diffusion process on time window data. The trained RNN generated signal reconstruction latent features in the form of a hidden state, which was then used in the reconstruction network to generate the final reconstructed signals. To detect faulty sensors, we compared the reconstructed signals with the original signals using residual analysis and judged signals with an error exceeding the fault threshold as faulty signals.

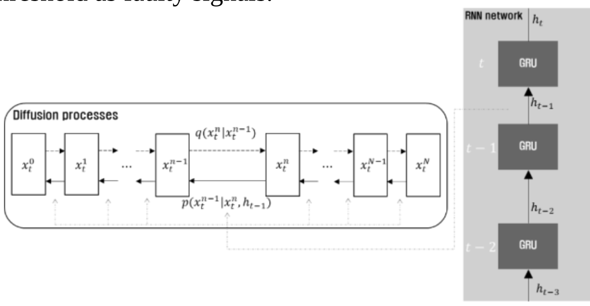

Fig. 2 Weight update on the RNN network from the denoising process

4. Results

4.1 Data description

We used a compact nuclear simulator (CNS) to create artificial intelligence training/test data for design basis accidents. The CNS is a 1D thermal-hydraulic code-based NPP simulator based on the Westinghouse 940Mwe pressurized water reactor, with simplified systems. We generated 960 training sets and 416 test sets for typical DBAs, selecting target sensors based on emergency operating procedures. Parameters monitored included PZR pressure and RCS average temperature, among others. Min-max normalization preprocessing was applied to the data. Refer to Table 1 for details.

Table I Dataset description

Accident type	No. of training set	No. Of test set
Loss of coolant accident	434	204
Steam generator tube rupture	104	31
Excess steam demand event	282	132
Loss of all feedwater	74	32
Reactor trip	66	17

Selected signal fault modes, including drift faults in upward/downward directions and stuck faults, were generated based on the literature on nuclear power plant cases and accidents. Fault scales were examined in two ranges, with faults injected 100 seconds after the reactor trip and single signal faults considered.

4.2 Training results with the diffusion model

The signal reconstruction model was evaluated using normal and faulty signal inputs. For normal signals, the reconstruction model showed accurate reconstruction of signals with stable performance. However, reconstructing signals for faulty data was found to be more challenging and resulted in instability. Faults with distinct trends, such as variables showing a downward trend during reactor shutdown and an upward drift fault caused by a fault effect, were detected with high accuracy. However, the reconstruction process tended to replicate the faulty signal when the fault type corresponded to the trend of the training data, resulting in a high rate of failure in detecting faults. To overcome this issue, enhancing the learning parameters were suggested.

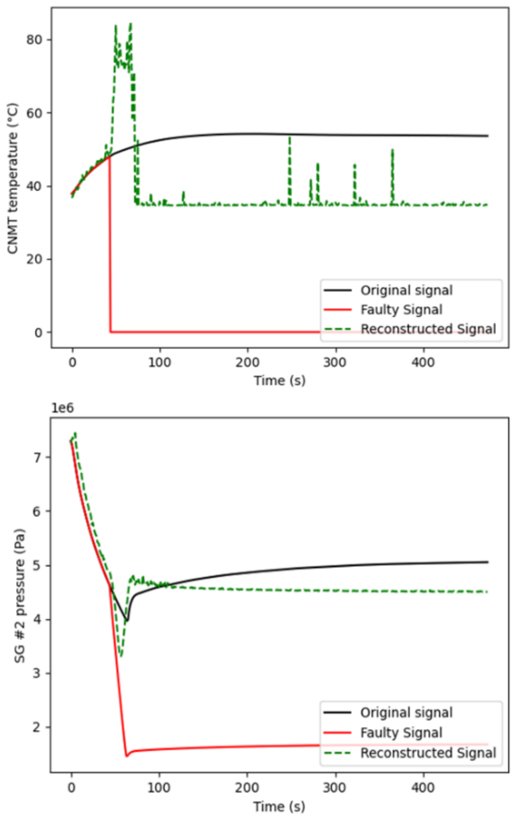

Fig. 3 Example of signal reconstruction with fault-injected data.

Fig. 3 showed the satisfying case of signal validation using a diffusion model with faulty data. The reconstructed signal successfully rebuilt the faulty signal with high residuals. The successful detection rate of fault data is 84.4% as table 2. The performance values of our proposed model are lower than those of the existing Aes, however, our model has excellence in the respect of nuisance error. In our previous research using autoencoders for signal reconstruction, nuisance errors occurred, resulting in normal signals being incorrectly identified as faulty signals at a rate exceeding 3%. In contrast, the diffusion model had a much lower incidence of nuisance errors as table 2, suggesting that it was more effective than the conventional autoencoder at detecting faults without generating a large number of false positives.

Table II Sensor fault detection accuracy of an autoencoder and diffusion model

			Diffusion model
Detection performance	Normal data	Fault detection rate	0%
	Fault data	Fault detection rate - Faulty sensor	84.4%
		Fault detection rate - Other normal sensors	0.003%

5. Conclusion

The diffusion model is an unsupervised artificial neural network that has shown promising results in signal validation for nuclear accidents. This study aimed to evaluate its robustness compared with AE, another widely used signal validation method. Our findings confirm that the diffusion model is more robust than AE, which enhances its applicability. However, further optimization of variables is necessary to improve fault detection rates and reduce model learning time. Combining the diffusion model with other methods, such as VAE, can enhance the accuracy and reliability of signal validation in nuclear accidents. Further research is needed to optimize the diffusion model's performance and explore its full potential in field applications. Our study shows the diffusion model's potential as a powerful tool for signal validation in nuclear accidents.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT). (No.RS-2022-00144042)

REFERENCES

[1] Kim, Jonghyun, Seungjun Lee, and Poong Hyun Seong. "Artificial Intelligence and Methods." Autonomous Nuclear Power Plants with Artificial Intelligence. Cham: Springer International Publishing, 2023. 9-28.

[2] Li, Xiangyu, et al. "Research on anomaly detection method of nuclear power plant operation state based on unsupervised deep generative model." Annals of Nuclear Energy 167 (2022): 108785.

[3] Li, Longyuan, et al. "Anomaly detection of time series with smoothness-inducing sequential variational

auto-encoder." IEEE transactions on neural networks and learning systems 32.3 (2020): 1177-1191.
[4] Dhariwal, Prafulla, and Alexander Nichol. "Diffusion models beat gans on image synthesis." Advances in Neural Information Processing Systems 34 (2021): 8780-8794.
[5] Yang, Ling, et al. "Diffusion models: A comprehensive survey of methods and applications." arXiv preprint arXiv:2209.00796 (2022).