Performance Evaluation in Anomaly Detection using Unsupervised Learning at Nuclear Power Plants

Sang Hyun Lee, Ji Hun Park, Ji Woo Hong, Man Gyun Na [*]
Department of Nuclear Engineering, Chosun Univ., 309 Pilmun-daero, Dong-gu, Gwangju, Korea 61452
[*]*Corresponding author: magyna@chosun.ac.kr*

1. Introduction

In nuclear power plants (NPPs), events occur due to various factors (i.e., equipment defects, human errors, etc.). When an event occurs, NPPs may enter a more serious situation if an operator does not take appropriate action. To prevent this situation, the operator needs to detect an anomaly quickly and take preemptive measures.

Many studies have recently been conducted on anomaly detection in NPPs using artificial intelligence (AI). In general, these studies use supervised and unsupervised learning among AI learning strategies. The supervised learning uses data with labels. However, the cost of the label is considered a disadvantage. In addition, labeling for NPPs has another problem. This is because there are fewer accidents at NPPs and there is little data for AI to learn. As a result, the unsupervised learning that does not require such labels is in the spotlight. This is because the unsupervised learning does not require labeling to learn data and solve problems. It also has the advantage of being useful for discovering data patterns that are generally not found.

In this paper, long short-term memory-autoencoder (LSTM-AE) and LSTM-variational autoencoder (LSTM-VAE) are used for the unsupervised learning-based anomaly detection. Additionally, performance evaluation will be conducted on a newly studied unsupervised anomaly detection (USAD). The proposed method is a time series-based method. This is considered appropriate for data from NPPs. This is because the data of NPPs are also classified as time-series data. Specifically, we will develop an anomaly detection model for a reactor coolant system (RCS) that is closely related to safety among various systems of NPPs. In addition, the performance evaluation of the developed model is performed to select the optimal model for anomaly detection in the NPPs system. Accuracy and F_1-score evaluation indicators are used to evaluate the performance of the developed model [1].

2. Methods

2.1 Long Short-Term Memory

LSTM is well known in the time series-based data method. LSTM is a method that overcomes the disadvantage of not having long-term dependencies to store old information and not being able to remember information that is far from recurrent neural network output [2].

LSTM has four characteristic layers. The first layer is the cell state. The cell state is divided into a short-term state h_t and a long-term state c_t. Second, the forget gate determines what information to be forgotten through the sigmoid layer in the forget gate. The forget gate is shown in Eq. (1).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

Third, the input gate determines which of the incoming information is to be stored in the cell state. After determining the information to be updated through the sigmoid layer, a new vector is created in the tanh layer. The input gate is shown in Eq. (2).

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

Finally, the output gate determines what information is to be output. After the output value is updated in the cell state, the same process is performed in the next cell. The output gate is shown in Eq. (3).

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{3}$$

In this paper, input data is used as time series data by combining the LSTM method with AE and VAE.

2.2 Autoencoder

AE is one of the unsupervised learning methods with the same number of neurons in the input and output layers. AE is characterized by symmetry starting with a latent variable located in the middle. AE consists of an encoder and a decoder. Fig. 1 shows the structure of AE [3]. The encoder compresses the input data. Compressed data goes through a decoder and is restored; here, the reconstructed data is not the same as the input data. This suggests that even if it is reconstructed well, some errors will exist. These errors are utilized in anomaly detection. The method to obtain reconstruction error (RE) is expressed as Eq. (4). The AE reconstruction error value is calculated by mean squared error, which is the difference between input and output.

$$RE(AE) = \left\| X - X' \right\|^2 \tag{4}$$

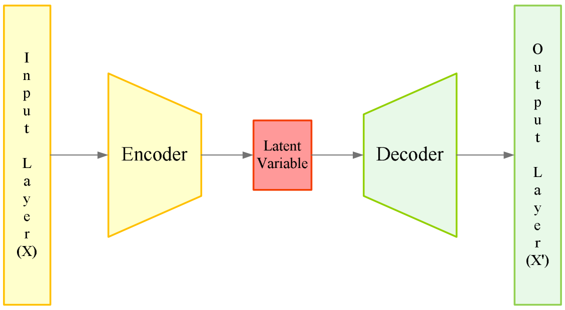

Fig. 1. The Structure of AE.

2.3 Variational Autoencoder

The VAE has a form similar to the AE, but the input data is emitted through the encoder as two outputs: average and standard deviation. Fig. 2 shows the structure of the VAE [4]. A Gaussian distribution is generated using the mean and standard deviation of the input data. This Gaussian distribution is generated using the encoder. In other words, unlike the previous AE method, VAE can learn the probability distribution for input data. The RE for error detection can be calculated in the same way as the RE of the preceding AE (refer to Eq. (4)).

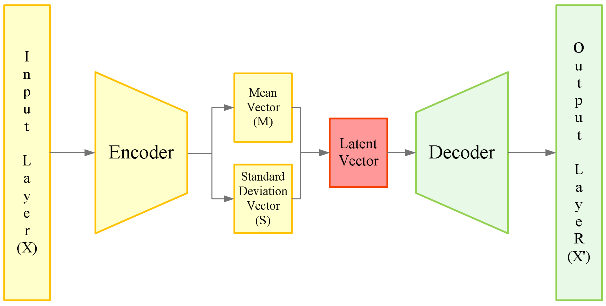

Fig. 2. The Structure of VAE.

2.4 Unsupervised Anomaly Detection

The USAD is one of the unsupervised learning methods specialized in multivariate time series data [5]. The USAD structure is a combination model that compensates for the disadvantage of the AE model and generative adversarial network (GAN). The disadvantage of AE is that it detects as normal when an outlier that is not much different from the threshold exists. The disadvantage of the GAN is learning instability. The AE model is a method that reconstructs input data. The GAN consists of a generator (G) and a discriminator (D). G reconstructs the input data similarly to the autoencoder model. D discriminates the reconstruction data output from G. In other words, G evaluates whether the reconstruction data is correctly reconstructed. If the reconstructed data is bad according to the evaluation result, G is retrained. Conversely, the reconstruction data derived from G is fed back to D. From this feedback process, G and D have a characteristic that their performances increase by complementing each other. This is referred to the adversarial structure. The RE expression for USAD is Eq. (5). Figs. 3 and 4 show the structure of the phase progression of the USAD.

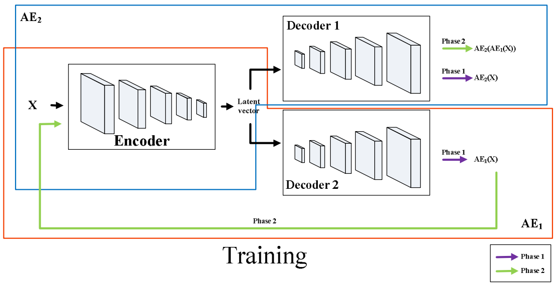

Fig. 3. The Structure of USAD phase 1.

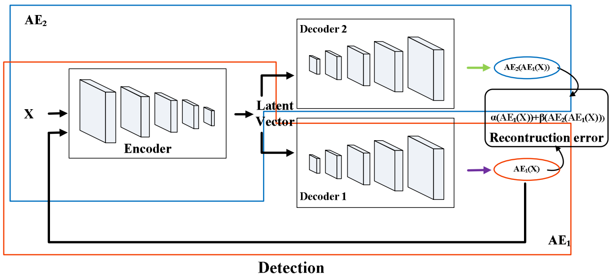

Fig. 4. The Structure of USAD phase 2.

$$RE(USAD) = \alpha(AE_1(X)) + \beta(AE_2(AE_1(X))) \qquad (5)$$

2.5 Anomaly Detection of Each Method

In anomaly detection using the unsupervised learning, each method learns normal data during learning. The input data is compared with the reconstructed data and it will be determined that the RE is normal if it is lower than the threshold, and abnormal if it is higher. Thresholds are calculated using the three-sigma theory widely used in the industry. So, only 99.7% of RE is considered normal, and other values are considered abnormal. The threshold is expressed using the mean (μ) and standard deviation (σ) of the RE. The threshold is shown in Eq. (6).

$$Threshold = \mu \pm 3\sigma \qquad (6)$$

3. Data processing

The CNS is a simulator designed with reference to the Westinghouse 993MWe Kori 3 and 4 NPPs. It can operate or monitor static and dynamic information through graphical representations. Through CNS, simulation data such as normal and abnormal state data of NPPs and equipment failure were obtained. Data were extracted from CNS, train data were normal data, and test data were data simulating abnormal scenarios.

Data preprocessing proceeds as follows: 1) train data augmentation 2) variable extraction, 3) data normalization, 4) data sliding window.

First, the amount of train data is insufficient, so noises are added to increase the amount of data. The reason for adding noise to train data is to solve the problem of insufficient train data for anomaly detection and to improve the size and quality of train data [6].

Second, there are a total of 2222 variables in the extracted data, and a total of 63 variables were selected by extracting the variables corresponding to the RCS system related to NPPs safety. The reason is to speed up the learning by reducing the size of the input variable. In addition, unnecessary variables can act as a degrading factor.

Third, the data were normalized. Normalization serves to find patterns by comparing the characteristics of the data. The reason for normalization is that certain characteristics can completely hide the characteristics of other data if there is a significant difference in the magnitude of the data. And it speeds up the training speed of the AI model. The min-max method is used for normalization, and it is shown in Eq. (7).

Finally, in this paper, all AI methods used are learned using time series data, and all extracted data are divided into 10 seconds using sliding window techniques. The reason is because empirically the performance was the best when divided into 10 seconds.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{7}$$

4. Result

In the anomaly detection based on the unsupervised learning, we used three methods: LSTM-AE, LSTM-VAE, and USAD. The data were divided into training and validation data, 90%, and 10%, respectively. Model optimization was performed by adjusting layers, the number of nodes, batch size, and time step. For AI model training, a total of 63 variables related to the RCS system were extracted, and data accumulate for 10 seconds using a sliding window technique in a time series-based method. In the model structure, because each method has the structure of an AE, which is a generative model, the number of nodes in the input layer is the same as the number of variables. So, the 63 extracted variables mentioned above are used as nodes in the input layer of each model. While AI training generally is faster as batch sizes get larger, AI training can become unstable. In order to compromise between learning speed and learning instability, batch size was empirically set to 64. And for the number of layers, the number of layers was optimized for each model. Hyperparameters were selected based on the results of each condition. The optimized model structure for each method is shown in Table I. Each model uses early stopping to prevent overfitting, which can result from excessive learning. However, underfitting occurs in an early stopping, each model sets the patience to 5 and waits for 5 epochs to stop the training even if there is no room for improvement in the validation loss. Threshold, which determines the presence or absence of abnormalities in anomaly detection, uses the three-sigma theory widely used in the industry. Fig. 5 shows anomaly detection graphs for three methods: Fig. 5(a) for the LSTM-AE method, Fig. 5(b) for the LSTM-VAE method, and Fig. 5(c) for the USAD method. In the figure, the blue line is the threshold that separates the anomaly from the normal, and the green dot is the part that the AI model detected to be normal as a result of the anomaly detection. On the other hand, the red dot is the part detected to be abnormal. In this paper, accuracy and F_1-score were adopted as indicators for evaluating the performance of the methods. Both evaluation values are indicators of how accurately the data is classified. These are calculated based on the confusion matrix as shown in Eqs. (8) and (9). The NPPs simulator data were composed of time series and USAD, a method for time series data, showed the best result with an accuracy of 0.966 and F_1-score of 0.981. The anomaly detection results of accuracy and F_1-score for each method are shown in Table II.

$$Accuracy = \frac{TN + TP}{TN + FP + FN + TP} \tag{8}$$

$$F_1\text{- }score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{9}$$

where

$$Precision = \frac{TP}{TP + FP}, \; Recall = \frac{TP}{TP + FN}$$

Table I: Optimized model structure for each method

Method	Layer	Node	Batch size	Time step
LSTM-AE	5	63	64	10
LSTM-VAE	9	63	64	10
USAD	5	63	64	10

Table II: The performance evaluation result of accuracy and F_1-score for each method

Method	Accuracy	F_1-score
LSTM-AE	0.911	0.925
LSTM-VAE	0.948	0.962
USAD	**0.966**	**0.981**

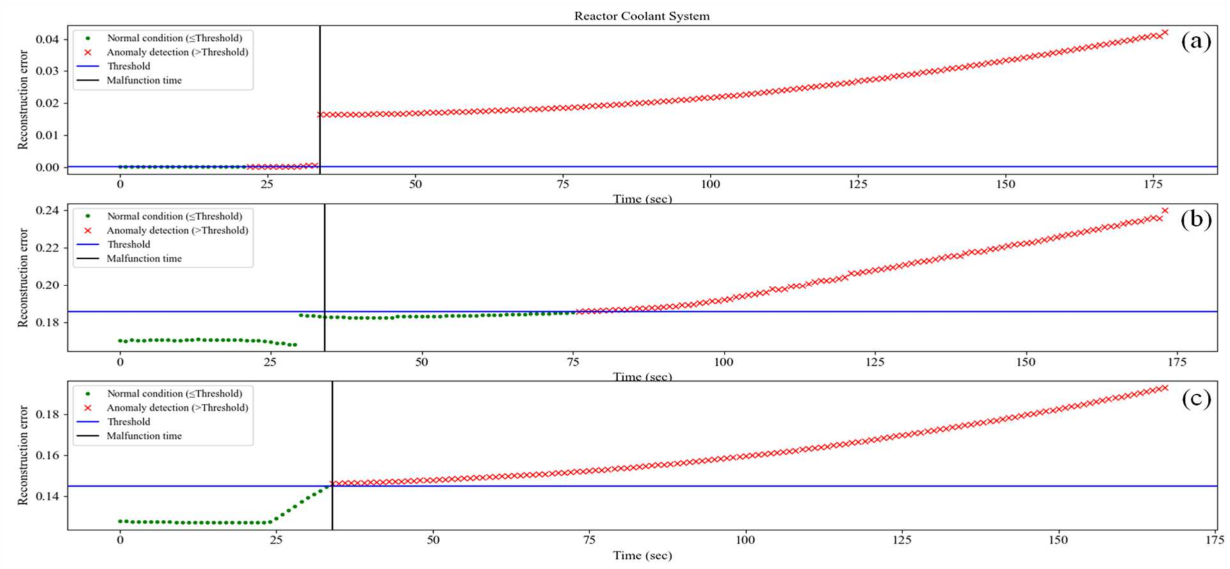

Fig. 5. Graph of anomaly detection result for each AI method. (a) LSTM-AE method, (b) LSTM-VAE method, (c) USAD method.

5. Conclusion

In this paper, an anomaly detection model was developed using NPP simulator data. Among time series-based unsupervised learning methods, LSTM-AE, LSTM-VAE, and USAD were considered to develop an anomaly detection model. Accuracy and F_1-Score are used as performance evaluation indicators to select the optimal anomaly detection model. Based on the performance evaluation results, the optimal anomaly detection model is USAD, and the accuracy and F_1-score are 0.966 and 0.981, respectively. It showed high performance compared to other compared methods. As a result, these research results are expected to be helpful in developing an anomaly detection model using time series-based unsupervised learning. In the future, the application of the VAE structure instead of the AE structure to the USAD model is considered. And the generative adversarial network technique which is a time series-based anomaly detection method will be applied.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant, funded by the Korean Government (MSIT) (Grant No. NRF-2018M2B2B106565123).

REFERENCES

[1] P. Niño, J. Omar, and F. Berzal, Evaluation metrics for unsupervised learning algorithms, arXiv preprint arXiv:1905.05667, 2019.
[2] Hochreiter. S, and Schmidhuber. J, Long short-term memory. Neural computation, Vol. 9, No. 8, pp. 1735-1780, 1997.
[3] M. Sakurada, T. Yairi, Anomaly detection using autoencoders with non-linear dimensionality reduction, in: Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis, ACM, p. 4, 2014.
[4] D. P. Kingma, S. Mohamed, D. J. Rezende and M. Welling, Semisupervised learning with deep generative models, In Advances in neural information processing systems, pp. 3581-3589, 2014.
[5] J. Audibert, P. Michiardi, F. Guyard, S. Marti, and M. A. Zuluaga., USAD: UnSupervised Anomaly Detection on multivariate time series, In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '20), pp. 3395–3404, 2020.
[6] Q. Wen, L. Sun, F. Yang, X. Song, J. Gao, X. Wang, and H. Xu, Time series data augmentation for deep learning: A survey, arXiv preprint arXiv:2002.12478. 2020.