Transactions of the Korean Nuclear Society Autumn Meeting
Changwon, Korea, October 20-21, 2022

Machine Learning Based Anomaly Detection for Quality Control of Spent Fuel Safety Information

Ga-Hee Sim, Moon-Ghu Park∗, Kyoon Ho Cha
Quantum and Nuclear Engineering, Sejong Univ., 209 Neungdong-to, Gwangjin-gu, Seoul, Republic of Korea
∗Corresponding author: mgpark@sejong.ac.kr

1. Introduction

According to the regulation for spent fuel transportation, when the consignor requests disposal, it is necessary to deliver spent fuel safety information, such as the type and characteristics, to the disposal facility operator [1]. Since the received spent fuel characteristic information directly affects the operation and safety of the disposal facility, the disposal facility operator needs to perform quality control on the data. Data quality control is advantageous in increasing the reliability of safety information, and anomaly detection is a representative method among quality control methods.

Anomaly detection is finding unexpected objects or data which differ from the norm. In an industrial field where anomaly detection is applied, there is a problem that normal data and abnormal data are unbalanced [2]. In this paper, we perform anomaly detection by models trained only with normal data, reflecting the difficulty of collecting abnormal data. A model trained only with normal data generates a small error for a normal pattern and a significant error for an abnormal pattern. If the data error exceeds the set threshold, the data is judged as abnormal.

Previous studies have used various methods for training with normal data, such as Support Vector Machine, Autoencoder, and Generative Adversarial Network [3, 4]. As another alternative, we propose a quality management system that detects anomalies using XGBoost. XGBoost is used in various fields with excellent performance, and it is known to prevent overfitting and be faster than the existing gradient boosting model [5]. In general, XGBoost is applied to problems with multiple input variables and one output variable, and this method is also applicable to multi-output regression and multi-label classification [6]. However, since the multiple output functions are supported experimentally, we combine several single-output XGBoost models and use them to predict multiple variables of spent fuel safety information.

2. Proposed Anomaly Detection Method for Quality Control

This section describes XGBoost used to predict data and presents an anomaly detection procedure for quality control of spent fuel safety information.

2.1. XGBoost Model

XGBoost is an algorithm that solves the gradient boosting problem of slow performance time and overfitting, as proposed by Chen and Guestrin [5]. XGBoost generates optimized models in a way that controls complexity to prevent overfitting while minimizing training loss. As a result, XGBoost's algorithm simplifies the inefficient search process and increases the model's computational efficiency and predictive power. XGBoost provides multiple output functions, but it is experimentally supported. Therefore, combining a single output model, as shown in Fig. 1, is necessary to predict multiple variables of spent fuel safety information.

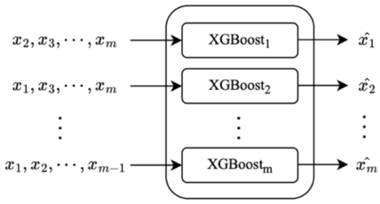

Fig. 1. Multiple XGBoost models.

2.2. Anomaly Detection Method

The anomaly detection method based on the multiple XGBoost models is as follows: 1) Implement steady-state estimation models for each variable using normal data of significant variables of spent nuclear fuel. 2) Predict each variable with the implemented estimation models. 3) Calculate the residual between the input data and predicted data. 4) If the sum of the residuals satisfies the anomaly detection condition, determine it as abnormal data. Fig. 2 shows the flowchart of anomaly detection.

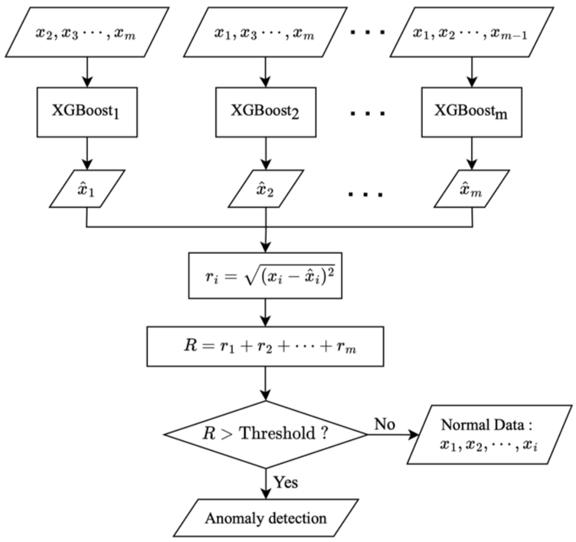

Fig. 2. Anomaly detection procedure.

3. Experiment Result

This section evaluates the accuracy of the implemented XGBoost for virtual spent fuel information and presents the proposed method's demonstration results.

3.1. Accuracy of XGBoost

Accuracy measures how correctly a model predicts the input values and is expressed as the mean squared error (MSE) between the observed values and predicted values, as shown in equation 1 [7].

$$A = \frac{1}{N}\sum_{i=1}^{N} (\hat{x}_i - x_i)^2 \qquad (1)$$

where $\hat{x}_i$ is the predicted value of the ith observation, x_i is the ith observation value, and N is the number of data. Table 1 shows the accuracy calculated in this experiment. Variable consists of release burnup, enrichment, and the total amount of U, U-235, Pu, Pu-239, and Pu-241 for spent fuel. Although v_3 and v_6 have lower accuracy than the other variables, all models can be considered very accurate.

Table. 1. Accuracy of auto-associative XGBoost.

Variable	*Accuracy*
v_1	3.511E-05
v_2	1.259E-05
v_3	0.00033
v_4	1.880E-05
v_5	1.820E-05
v_6	0.00028
v_7	1.986E-06

3.2. Anomaly Detection

We experimentally validate the proposed method by including the fault input in v_1. The fault inputs are randomly generated and contain values similar to the original values, and Fig. 3 shows the difference between the original values and the fault values. The threshold of the experiment is the 95% confidence interval for the residual between the observed and predicted values. Table 2 summarizes the anomaly detection results for the fault inputs. Indexes classified as 'normal' show little difference between the original and fault values, and the indexes classified as 'anomaly' show a large difference. This result means that the proposed method accurately detects anomalies.

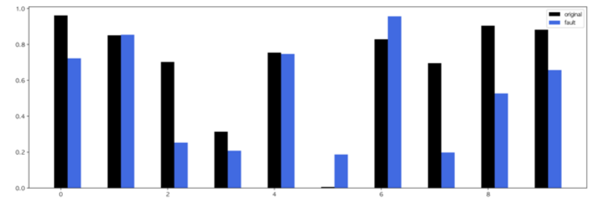

Fig. 3. Comparison of original data and defect data.

Table. 2. Result of anomaly detection (threshold=0.0359).

Index	*Sum of residuals*	*Result*
0	0.444	Anomaly
1	0.021	Normal
2	1.246	Anomaly
3	0.724	Anomaly
4	0.028	Normal
5	0.949	Anomaly
6	0.143	Anomaly
7	1.422	Anomaly
8	0.704	Anomaly
9	0.404	Anomaly

4. Conclusion

We propose an anomaly detection method based on XGBoost for quality control of spent fuel safety information. As a result of performing the anomaly detection by the proposed method, the XGBoost model showed high accuracy in the normal inputs and classified the anomaly by providing a stable prediction even in the abnormal inputs. In future studies, we plan to optimize the hyper-parameters of XGBoost so that the accuracy of all variables is similar and identify the variables with anomalies by comparing the residuals.

REFERENCES

[1] 원자력안전위원회, 사용후핵연료 인도규정, 원자력안전위원회고시 제 2021-22 호.
[2] R. Chalapathy, S. Chawla, Deep learning for anomaly detection: A survey, 2019.
[3] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, H. Chen, Deep autoencoding gaussian mixture model for unsupervised anomaly detection, in: International Conference on Learning Representations, 2018.
[4] D. Li, D. Chen, J. Goh, S.-k. Ng, Anomaly detection with generative adversarial networks for multivariate time series, 2018.
[5] T. Chen, C. Guestrin, XGBoost: A Scalable Tree Boosting System, Association for Computing Machinery, 2016.
[6] X. developer, xgboost : Release 2.0.0-dev, 2022.
[7] J. Hines, D. Garvey, R. Seibert, A. Usynin, Technical Review of On-Line Monitoring Techniques for Performance Assessment, Vol. 2, US Nuclear Regulatory Commission, 2007.