4th Asia Pacific Conference of the Prognostics and Health Management,
Tokyo, Japan, September 11 – 14, 2023 R02-10

Abnormal Detection Using Two-Stage Method in Combined Power Plant

Yun hee Kim[1], Sang Woo Ha[2], Sung shim Kim[3], Moon won Lim[4], Sung Yong Chang[5],

Hae Su Kang[6], Gyu Sang Park[7], Mi Lee Joo[8] and Suk Joo Bae[9†]

[1,2,3,4,9]*Department of Industrial Engineering, Hanyang University, Seoul, Korea*

[5,6,7,8] *Korea Electric Power Corporation (KEPCO), Daejeon, Korea*

yuuun@psm.hanyang.ac.kr

[†]*sjbae@hanyang.ac.kr*

ABSTRACT

Complex systems, such as power plants, demand precise and reliable anomaly detection mechanisms. Traditional supervised learning approaches often fall short due to the challenges of imbalanced data and the scarcity of labeled abnormal instances. This paper introduces a two-stage methodology to address these challenges. The first stage emphasizes feature engineering, mitigating redundant sensor effects, and reducing dimensionality through Kmeans clustering and PCA. The second stage employs an LSTM-Autoencoder for abnormal event detection. Validated using data from a combined power plant, our approach demonstrates superior performance over existing techniques in terms of accuracy and computational efficiency. This research not only advances the field of anomaly detection in power plants but also offers insights for other complex systems.

1. INTRODUCTION

Complex systems, exemplified by power plants, necessitate the reliable and precise identification of abnormal events. The inherent intricacies of such systems, compounded by the vast array of sensors, variable system usage, and constrained computational resources, make anomaly detection a formidable task. Typically, the data is imbalanced, with a vast majority of normal operational data and a sparse representation of abnormal scenarios. Furthermore, obtaining labeled data for abnormal situations is difficult, making traditional supervised learning approaches less feasible. While the literature is replete with diverse strategies ranging from probabilistic models to machine learning and deep learning techniques, there remains a pressing need for more efficient and accurate methods. In this context, we introduce a novel two-stage methodology. Initially, we focus on feature engineering, aiming to mitigate the effects of redundant sensors and to condense the sensor dimensionality through clustering and Principal Component Analysis (PCA). Subsequently, we employ an Autoencoder for the pivotal task of abnormal event detection. Our methodology is predicated on the principle of leveraging a select subset of sensors, those that manifest significant variability and are crucial for anomaly detection.

To validate the efficacy of our approach, we harness data from an operational combined power plant. Empirical results underscore the superiority of our method over extant techniques, both in terms of detection accuracy and computational efficiency. Notably, our strategy holds the promise of substantially curtailing the computational overhead associated with anomaly detection in intricate systems, rendering it apt for practical implementations.

2. RELATED WORK

An overview of anomaly detection techniques reveals two main categories: conventional and data-driven techniques. Conventional techniques draw from established fields such as statistics and signal processing. Data-driven techniques, on the other hand, rely on machine learning and data mining.(L Erhan, M Ndubuaku, M Di Mauro, W Song, &M Chen, 2021)

In the data-driven approach, many works have been proposed to detect abnormalities, leveraging the advancements in machine learning techniques. These include proximity-based classifiers(Mani, &Zhang, 2003), support vector machines (SVM)(Yin, Yang, &Pan, 2008), and spatio-temporal autoencoders(Chong, &Tay, 2017).

3. METHODOLOGY

The process for detect abnormal in combined power plant using two-stage approach is shown in the table 1 below.

Table 1. Process for Methodology

Stage #1 Feature Engineering	- Mahalanobis distance for eliminating redundant sensors - Kmeans clustering for grouping similar data points - PCA for dimensionality reduction
Stage #2 Abnormal Detection	- LSTM-Autoencoder for anomaly detection based on reconstruction error

In the methodology, initial stage focuses on processing the raw data to extract meaningful features and reduce dimensionality. And then next stage involves applying a Autoencoder to identify abnormalities based on the features derived in the first stage.

3.1. Stage One : Feature Engineering

In the Stage One, The primary objective of this stage is to process the raw data, extract meaningful features, and reduce dimensionality. This ensures that the data is transformed into a format that is more suitable for the subsequent abnormal detection phase.

3.1.1. Mahalanobis Distance

To account for the effect of multiple similar sensors, we compute the Mahalanobis distance between sensors. Unlike the Euclidean distance, the Mahalanobis distance considers the covariance of the data, making it more effective in identifying outliers and eliminating redundant sensor trends.

3.1.2. Kmeans Clustering

After addressing the influence of redundant sensors, we utilize Kmeans clustering to group similar data points. Kmeans clustering is a partitioning method that divides the dataset into a set of mutually exclusive clusters. By doing so, it helps in discerning the inherent groupings within the data. This step not only aids in understanding the underlying structure of the data but also sets the stage for subsequent dimensionality reduction.

3.1.3. PCA for Dimensionality Reduction

Principal Component Analysis (PCA) is applied post clustering to condense the data's dimensionality within each cluster. By retaining only the most significant features, we ensure that the subsequent abnormal detection phase is both efficient and effective.

3.2. Stage Two: Abnormal Detection

With the features engineered in the first stage, we now focus on the pivotal task of identifying abnormalities

3.2.1. LSTM-Autoencoder

The LSTM-Autoencoder is a specialized neural network architecture that combines the strengths of Long Short-Term Memory (LSTM) networks with the principles of autoencoders. Trained on normal operational data, the LSTM-Autoencoder's reconstruction error serves as a metric for detecting abnormalities. A higher reconstruction error typically indicates potential anomalies, allowing for timely interventions in the power plant's operations.

4. EXPERIMENT AND ANALYSIS

4.1. Data Description

The foundation of our study is a dataset sourced from a Combined Power Plant, generously provided by the KEPCO Research Institute (KEPRI). This dataset is invaluable in offering insights into the intricate operations of power plants and serves as the testing ground for our proposed methodology.

The dataset captures a broad spectrum of sensor readings, reflecting the diverse operations of the Combined Power Plant. These readings, collated under various operational scenarios, are presented in Figure 2.

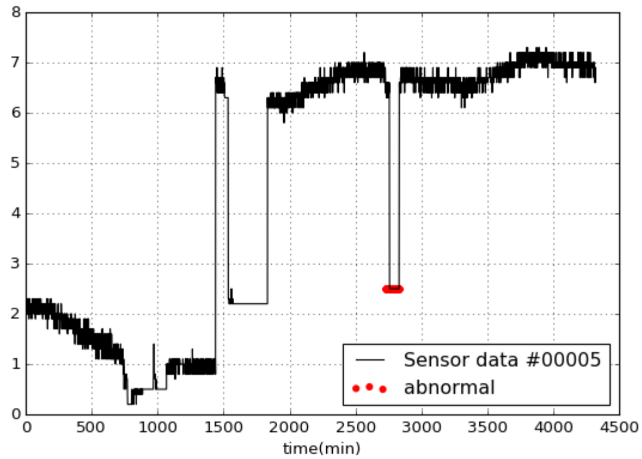

Figure 1. Sensor data and abnormal time

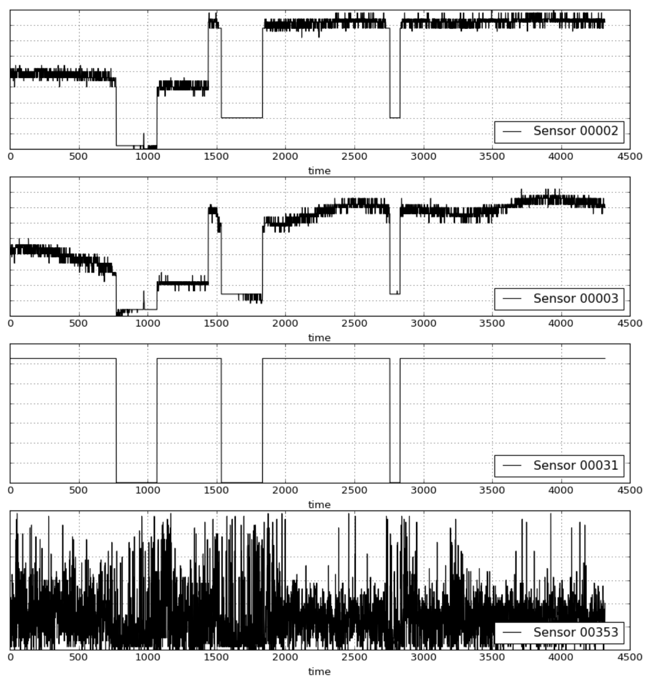

Figure 2. Sensor Data Deviation: A Comparative Overview

4.2. Feature Engineering Result

By calculating the Mahalanobis distance between sensors, we can gauge their similarity. A distance value close to 0 indicates high similarity between the sensors, implying that their readings are often in sync or exhibit similar patterns. This interpretation is visually represented in Figure 3, which showcases sensors with high similarity based on their Mahalanobis distance.

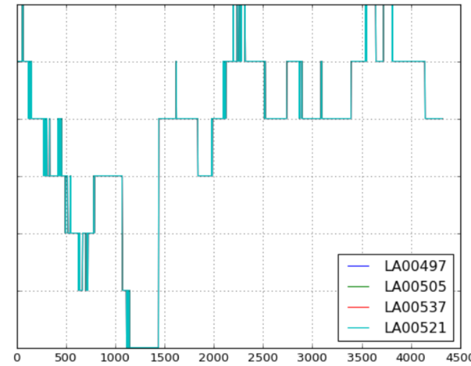

Figure 3. Example of high similarity sensors

Upon evaluating sensor similarities, we employed Kmeans clustering to categorize sensors based on their inherent patterns and inter-relationships. Kmeans clustering segregates the dataset into distinct clusters, aiming to group data points that are more similar to each other than to those in other clusters.

To ascertain the optimal number of clusters, we utilized the silhouette score. This score measures how similar an object is to its own cluster compared to other clusters. Through rigorous analysis, it was observed that the silhouette score reached its maximum value when using two clusters. This indicates that two clusters represent the most suitable configuration for our dataset.

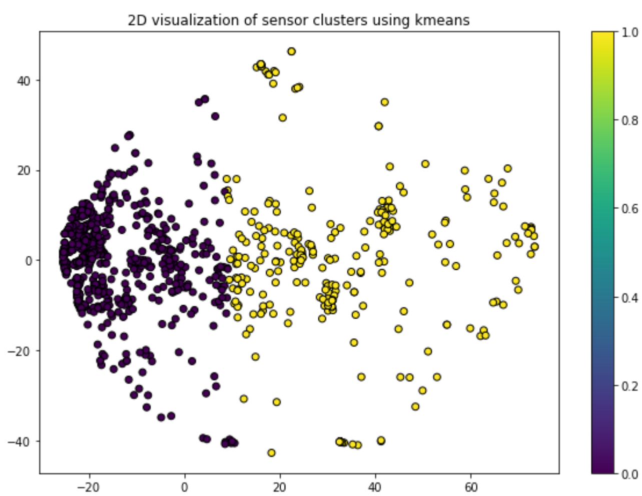

Figure 4. Scatter plot of kmeans cluster

Subsequent to the clustering process, we embarked on a dimensionality reduction phase for each cluster using Principal Component Analysis (PCA). The objective was to retain the most informative features while reducing computational complexity. Our criterion for the number of principal components to retain was set by the explained variance ratio, aiming for a cumulative explained variance of

0.80%. This ensures that the reduced dataset maintains 80% of the original data's variance, striking a balance between data simplification and information retention.

4.3. Abnormal Detection Result

In the pursuit of robust anomaly detection, we employed the LSTM-Autoencoder, a deep learning model tailored for time-series data. This model was predominantly trained on data representing normal operational conditions. Given its architecture, the LSTM-Autoencoder excels at capturing long-term dependencies and intricate patterns inherent in time-series data. This capability is especially pertinent for our application, considering the pronounced fluctuations present in our dataset.

To ascertain the most effective model configuration, the proposed model was compared with LSTM-Autoencoder model and Spectral Clustering with the LSTM-Autoencoder. The rationale behind this comparative analysis stemmed from the high correlation observed within our data. While dimensionality reduction techniques, such as Principal Component Analysis (PCA), can simplify the data structure, they may inadvertently eliminate critical information. As such, clustering methods, particularly Spectral Clustering, offer an alternative approach to discern patterns without compromising data integrity.

For a comprehensive assessment of our anomaly detection model, we employed four key metrics: True Alarm Rate (TAR), Missed and false Alarm Rate (M&FAR), True and false Alarm Rate (T&FAR), and False Alarm Rate (FAR). These metrics provide a holistic view of the model's capability to accurately detect anomalies while minimizing false alarms.

The proposed model shows good performance in detecting abnormal instances in the system, which can be used to identify utility potential in real-world applications.

	TAR	M&FAR	T&FAR	FAR
LSTM-Autoencoder	100	25.73	31.50	27.30
Kmeans+ LSTM-Autoencoder	100	21.94	27.71	23.28
Spectral+ LSTM-Autoencoder	100	26.72	32.49	28.35

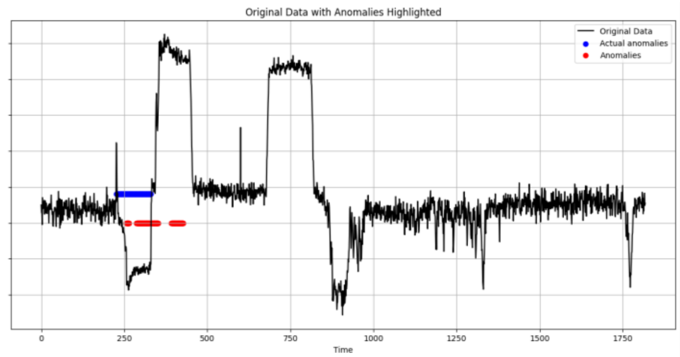

Figure 5 Original Data and Anormal detection result

5. CONCLUSION AND FUTURE WORK

In this study, we addressed the challenge of anomaly detection in complex systems, specifically power plants, by introducing a novel two-stage methodology. The first stage focused on feature engineering, where redundant sensors were addressed, and dimensionality was reduced using techniques like Mahalanobis distance calculation, Kmeans clustering, and PCA. The second stage centered on the application of LSTM-Autoencoder for abnormal event detection. Our empirical results, derived from a dataset provided by KEPCO, demonstrated the efficacy of our approach, outperforming traditional methods in both accuracy and computational efficiency. The proposed model not only offers a robust solution for anomaly detection in power plants but also holds significant promise for other complex systems, emphasizing its potential for broader real-world applications.

ACKNOWLEDGEMENT

This work was partly supported by Korea Institute of Energy Technology Evaluation and Planing(KETEP) grant funded by the Korea government(MOTIE) (20217010100020, Development of Intelligent O&M Technologies for Standard Combined Cycle Power Plant)

REFERENCES

Erhan, L., Ndubuaku, M., Di Mauro, M., Song, W., Chen, M., Fortino, G., ... & Liotta, A. (2021). Smart anomaly detection in sensor systems: A multi-perspective review. Information Fusion, 67, 64-79.

Mani, I., & Zhang, I. (2003, August). kNN approach to unbalanced data distributions: a case study involving information extraction. In Proceedings of workshop on learning from imbalanced datasets (Vol. 126, pp. 1-7). ICML.

Yin, J., Yang, Q., & Pan, J. J. (2008). Sensor-based abnormal human-activity detection. IEEE transactions on knowledge and data engineering, 20(8), 1082-1090.

Chong, Y. S., & Tay, Y. H. (2017). Abnormal event detection in videos using spatiotemporal autoencoder. In Advances in Neural Networks-ISNN 2017: 14th International Symposium, ISNN 2017, Sapporo, Hakodate, and Muroran, Hokkaido, Japan, June 21–26, 2017, Proceedings, Part II 14 (pp. 189-196). Springer International Publishing.