NIPPON STEEL TECHNICAL REPORT No. 131 OCTOBER 2024

Technical Report

Unsupervised Anomaly Detection in Mixed Processes Using Clusters

Masataka INOUE*

Abstract

In this paper, we introduce Polygon-AI as a unique technique for detecting anomalies in operations and equipment of the manufacturing process. In the steel industry, which is based on high-mix, small-lot, mass continuous production, there is a demand for anomaly detection technologies that can deal with a wide variety of operation modes. Polygon-AI applies non-negative matrix factorization, a data clustering method to detect anomalies. It enables us to automatically extract a normal model that describes multiple linear relations between variables in time-series data and quantify the degree of abnormality by gauging the distance between current data and the normal model. Currently, Polygon-AI is mainly applied to power generation and energy facilities. This paper explains the theory of Polygon-AI and introduces its practical applications.

1. Introduction

The steel industry is a material and process industry that operates continuously 24 hours a day and 365 days a year. To maintain competitiveness, high quality, low cost, and mass production are essential. The priority is to maintain and continue production by stabilizing operations and equipment. In particular, power generation and energy plants face more stringent requirements from upstream processes such as blast furnaces and steel plants. Once an anomaly or problem occurs and shuts down operations and equipment for a long period, it leads to a loss of production opportunities at the entire steelworks. It is important to detect the occurrence of anomalies as early as possible and remove their cause. Against this background, the steel industry has a high need for detection of operation and equipment anomalies and has made various efforts from the past to the present.

The most common anomaly detection method is to set an alarm threshold in the time series data of monitoring target operation and equipment and to issue an alarm when the alarm threshold is reached. This method is effective if operators and maintenance workers who are familiar with the operation and equipment can set appropriate alarm thresholds for anomalies. However, the steel industry as a process industry has enormous types and numbers of monitoring targets and must set alarm thresholds for specific monitoring targets. The management of the alarm thresholds requires enormous effort. Experienced workers may be able to detect an anomaly precursor before the alarm threshold is reached, but it is difficult for humans to comprehensively and continuously monitor anomalies.

On the other hand, if the mechanism of anomaly occurrence is known, it becomes possible to detect anomalies based on statistical and physical models. However, the use of such models is limited because there are only a limited number of cases in which it is possible to identify the mechanism of anomaly occurrence. When equipment and processes are large and complex, it becomes difficult to construct a deductive model. In this sense, the use of deductive models is limited.

Against this background, in recent years, it has become common to apply machine learning to construct anomaly detection models.[1] Machine learning does not require the identification of anomaly occurrence mechanisms in advance in order to build a model. Instead, it requires time series data observed in the past as input. Generally, machine learning algorithms used for anomaly detection are unsupervised learning (learning that automatically extracts data characteristics). The time series data (normal data) of operation under normal conditions are input. The reasons for this are that it is difficult to collect large amounts of operation and equipment anomaly data and

* Senior Manager, Control Development Section, Systems & Control Technology Dept., Systems & Control Engineering Div., Plant Engineering and Facility Management Center
20-1 Shintomi, Futtsu City, Chiba Pref. 293-8511

that it is more rational to judge "not normal" than to judge an anomaly.[2)]

However, simply applying machine learning to data that includes a wide variety of operation and operation modes may yield less reliable models. One way to solve these problems is to build machine learning models for the respective operation modes and then switch the models with the operation modes. However, building a model requires humans to sort data into appropriately sized sets by trial and error. This requires enormous effort.[3)]

On the other hand, if advanced machine learning and AI are applied to obtain a highly reliable model, a black box model can be obtained. Anomaly detection often requires that the obtained results must be capable of being interpreted and explained. The model should be as simple as possible in this sense.

Considering the above-mentioned anomaly detection technology issues and needs, we independently developed Polygon-AI, an anomaly detection technology that applies data clustering. Data clustering is a method of sorting similar data into subsets or clusters from a large amount of data. It is empirically known that a linear relationship exists between variables in time series data obtained from processes.[4, 5)] Polygon-AI is Nippon Steel's proprietary technology that automatically extracts normal models that express multiple linear relationships contained in the time series data and calculates the distance between the time series data and the normal model at each point of time as the anomaly score. In the chapters that follow, we describe the theory and application examples of Polygon-AI.

2. Anomaly Detection in Operation and Equipment with Polygon-AI

First, we describe non-negative matrix factorization (NMF). NMF is one of the data clustering methods and forms the basis of Polygon-AI. Next, we describe how time series data consisting of multiple variables observed in the past is treated as a matrix and that common characteristic factors (basis matrices) contained in the matrix are extracted by applying NMF.

Polygon-AI approximates the time series data at respective points in time by multiplying the extracted basis matrices by coefficient matrices. The spaces expressed by the products of the basis and coefficient matrices are plausible time series data at respective points in time. The plausible time series data are estimated from the time series data observed in the past. The distance between the plausible time series data and the actual time series data is calculated as the anomaly score.

2.1 Non-negative matrix factorization

Non-negative matrix factorization (NMF) is a method that decomposes the original matrix into a product of matrices, all of which contain only non-negative values. NMF is a data clustering method. It extracts the characteristic factors (base matrix) of the original matrix by imposing the condition that it does not contain negative values. The results are expressed as the addition of the characteristic factors.

$$Y \sim \Phi \cdot X \tag{1}$$

where Y is the original matrix, Φ is a coefficient matrix, and X is a basis matrix. All are non-negative real matrices. NMF decomposes a matrix into the product of a coefficient matrix and a basis matrix. The elements of the resulting coefficient matrix have the property of being mathematically sparse or mostly 0.

It is generally known that NMF solutions are not uniquely determined. NMF is defined as the optimization problem shown below.

$$\begin{aligned} & Y \sim \Phi \cdot X \\ & \text{minimize } \ \|Y - \Phi X\|_F \\ & \text{subject to} \\ & Y \in R^{T \times M},\ \Phi \in R^{T \times N},\ X \in R^{N \times M} \end{aligned} \tag{2}$$

where $\|\cdot\|_F$ represents the Frobenius norm (the square root of the sum of squares of each matrix component). Y, Φ, and X are non-negative real matrices. Y has a size of T rows and M columns, Φ has a size of T rows and N columns, and X has a size of N rows and M columns. As a method for solving the optimization problem of Eq. (2), an efficient updating algorithm is proposed by Lee and Seung.[6)] The specific procedure is as follows:

1. Give random non-negative values as the initial values of Φ and X.
2. Solve the minimization problem under fixed Φ and update X.
3. Solve the minimization problem under fixed X and update Φ.
4. Repeat the iterative calculations 2 and 3.

The above NMF optimization problem becomes a local solution because the solution X changes with the initial values. Therefore, such an approach is generally adopted whereby calculations are performed with various initial values and a solution close to the global solution is selected from multiple solutions.[7)]

2.2 Polygon-AI

Similar to common machine learning methodologies, Polygon-AI consists of an algorithm that learns a normal model for anomaly detection and an algorithm that applies the learned model to actual time series data.

2.2.1 Learning algorithm

In this section, we describe the learning algorithm. Polygon-AI treats previously observed time series data as a matrix and applies NMF to the matrix. However, since NMF imposes a non-negative condition on the matrices, it cannot be directly applied to the process data time series that assume various values. Therefore, we perform preprocessing to normalize the time series data to (0, 1) by rescaling them in the measurement range of each variable.

Let us consider Eq. (2) again. Y is the time series data observed in the past, Φ is a coefficient matrix, X is a basis matrix, T is the number of samples in the time series data, and M is the number of variables (or items). Polygon-AI applies NMF to Y and sorts Y into clusters with similar modes. The basis matrix X calculated at this time is a "matrix corresponding to the common characteristics of all samples of operation and operation modes." It is referred to as the operation state. The row size N of X is the number of clusters. **Figure 1** shows an image of the non-negative matrix factorization in Polygon-AI.

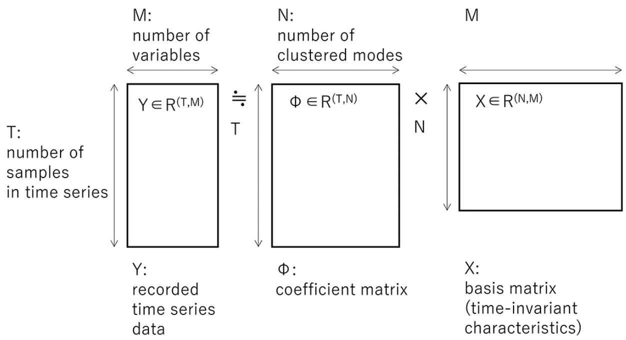

Fig. 1 Formulation of non-negative matrix factorization in Polygon-AI

However, it isn't easy to apply NMF to the time series data directly. This is because although the Frobenius norm minimizes the error of the entire matrix, it does not guarantee the restoration of the row data. This means that the linear relationship that is invariably present between the variables at respective points in time is not maintained. Therefore, Polygon-AI attempts to restore as many samples of the previously observed time series data as possible by devising the evaluation function of the optimization problem in NMF. First, the approximation error e is defined as the Euclidean norm of each row (the square root of the sum of squares of each vector component).

$$\begin{aligned} &D = Y - \Phi \cdot X \\ &D = (d_1\ d_2\ \dots\ d_T)^T \\ &D \in R^{T\times M},\ d_i \in R^{1\times M} \\ &e = (|d_1|, |d_2|, \dots, |d_T|)^T \end{aligned} \tag{3}$$

Suppose the simple Euclidean norm of e is used as the evaluation function. In that case, it may lead to an overestimation due to local errors (such as the worst value of $|d_i|$) in previously observed data at a specific time point. Increasing the number of clusters N can be an effective strategy to mitigate such local errors. However, if N is set to be too large, it can result in insufficient data for each cluster and compromise statistical reliability.

Generally, NMF calculations assume that the number of clusters N is known.[8] However, if the number of actual operations and operation modes is used as the known N, the number of clusters becomes too large. On the other hand, it is not easy for humans to determine N through trial and error. Therefore, Polygon-AI attempts to find X that most accurately approximates Y with the minimum N by incorporating N into the iterative calculations to solve the optimization problem.

Specifically, the number of times β (number of rows) that the Euclidean norm $|d_i|$ of each row of the approximation error e exceeds the threshold is defined as an evaluation function. Then, the threshold of $|d_i|$ is defined by Eq. (4) as the maximum value in the box-whisker plot (**Fig. 2**).

$$\text{Threshold} = \text{Third quartile} + \frac{(\text{Third quartile} - \text{First quartile})}{2} \tag{4}$$

when $|d_i|$ is arranged in ascending order, the data at the 1/4 from the bottom is defined as the first quartile, and the data at 3/4 from the bottom is defined as the third quartile. Generally, anomaly detection assumes that Y is normal data, but actual process data inevitably contains outliers (noise). Therefore, by excluding outliers in approximation errors, the sensitivity of X to outliers in process data is reduced.

In this way, Polygon-AI is characteristic in that it is formulated to remove outliers from samples of time series data observed in the past, restore the data as much as possible, and automatically derive just enough clusters. This is based on the perspective of engineers who want to build simple and highly reliable anomaly detection models that can handle time series data, including various operations and operation modes.

As mentioned above, in NMF, the solutions Φ and X change depending on the initial values of the coefficient matrix Φ and the basis matrix X. Therefore, in the learning algorithm of Polygon-AI, appropriate initial values are set for Φ and X, and NMF is repeatedly performed. Then, by repeating NMF while changing N, we find the minimum N, X, and threshold that minimize β.

2.2.2 Application algorithm

This section describes the application algorithm of Polygon-AI, which is the procedure for applying it to actual time series data and judging that the actual time series data are "not normal." The operation state (X) obtained from the learning algorithm described above represents a set of all time series data that can be estimated from the previously observed time series data Y. This corresponds to the fact that the normal model of Polygon-AI is a space spanned by the basis vectors x_i that compose the operation state X.

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \qquad X \in R^{N\times M},\ x_i \in R^{1\times M} \tag{5}$$

For the time series data y observed online at a certain point in time, the anomaly score and deviation score concerning the normal model of Polygon-AI are calculated as follows:

$$\begin{aligned} &y \sim \phi X \\ &\text{minimize } |y - \phi X| \\ &\text{subject to} \\ &y \in R^{1\times M},\ \phi \in R^{1\times N},\ X \in R^{N\times M} \end{aligned} \tag{6}$$

where ϕ is a coefficient vector calculated using NMF under non-negative constraints. The Euclidean norm of the estimation error for ϕ obtained by Eq. (6) is defined as the anomaly score. In addition, the estimation error is defined as the deviation score.

$$\text{anomaly score} = |y - \phi X| \tag{7}$$

$$\text{deviation score} = y - \phi X \tag{8}$$

Each element of the deviation score represents the contribution score of a variable that causes the time series data y to be "not normal" at a certain point in time. Furthermore, the index of the largest element in the ϕ array corresponds to the mode.

$$\text{mode} = arg\ max\ (\phi) \tag{9}$$

The above is the theory of Polygon-AI. **Figure 3** schematically

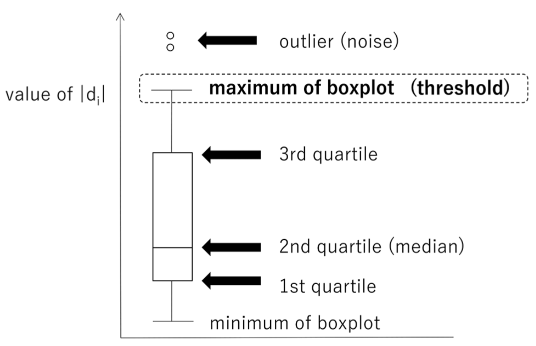

Fig. 2 Threshold for approximation error in Polygon-AI

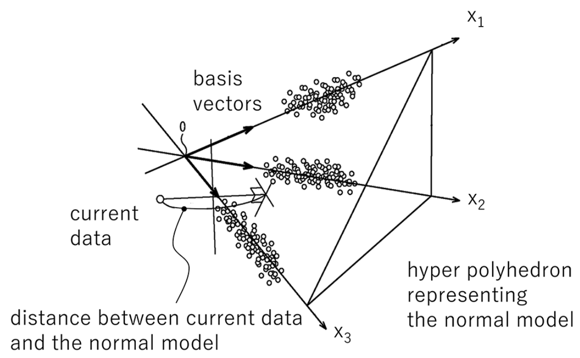

Fig. 3 Schematic diagram of anomaly detection by Polygon-AI

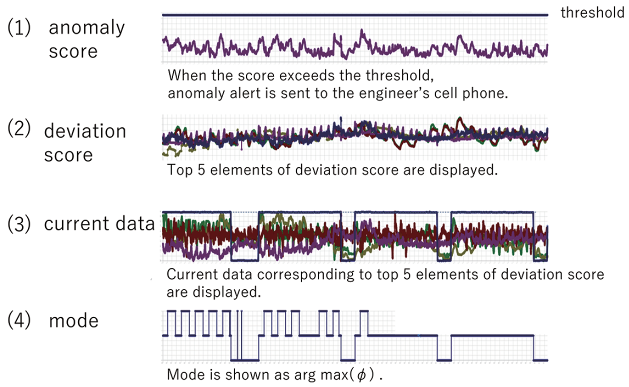

Fig. 4 Example of monitoring screen of Polygon-AI system

illustrates anomaly detection using Polygon-AI. The distance between the current data and the plausible data estimated from the space spanned by the basis vectors x_i representing the operation state is the anomaly score. This technology was named Polygon-AI because it automatically derives a polygon that can accurately explain the multiple linear relationships contained in the multidimensional time series data.

Figure 4 shows an example of a monitoring screen of the anomaly detection system using Polygon-AI. The first row shows an anomaly score trend. The anomaly score is automatically emailed to the engineer when it exceeds the threshold. The second row shows the deviation score with respect to the anomaly score. The top five variables with the highest contribution score of the "not normal" factors are automatically extracted to support the cause analysis of operation and equipment anomalies.

3. Application Examples

In this chapter, we introduce specific application examples of Polygon-AI.

3.1 Needs for anomaly detection in power generation and energy facilities

Power generation and energy facilities at the steelworks are responsible for supplying electricity to respective plants at the steelworks and to surrounding areas. Any operation or equipment problem can have a significant impact on the production of the steelworks and the lives of residents. By detecting early signs of operation and equipment anomalies, it becomes possible to avoid penalties to the local power company due to insufficient power generation at the steelworks, prevent the spread of secondary damage, and reduce downtime. Moreover, it is expected to ease the monitoring work of operators and alleviate the anomaly identification work of maintenance personnel.

3.2 Boiler tube leaks

A boiler tube leak is a phenomenon where the boiler tube is damaged due to aging or thermal fatigue. Once it occurs, there is a possibility of secondary damage from erosion from steam blowing out of the leak (**Fig. 5**). In addition to the penalty to the power company for insufficient power generation, it requires a significant amount of effort to locate the leak. There is a high need for anomaly sign detection and anomalous point identification.

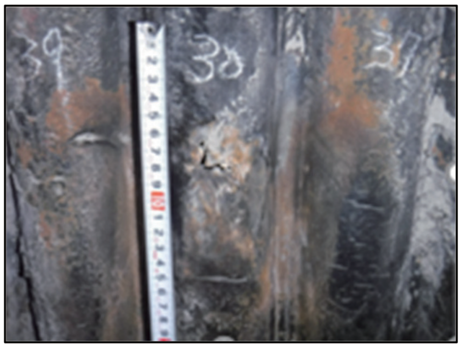

Fig. 5 Appearance of boiler tube rupture

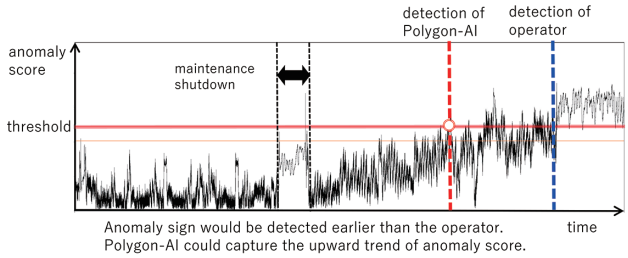

Fig. 6 Polygon-AI validation result for the boiler tube leak incident

3.3 Polygon-AI application verification

Figure 6 shows the analysis results of Polygon-AI applied to the time series data of a boiler tube leak that actually occurred at one of the power generation and energy facilities at Nippon Steel.

Polygon-AI detected an anomaly sign earlier than the detection of the boiler tube leak by the operator. It has also been confirmed that the variables that correspond to an increase in the deviation score do not contradict equipment and process knowledge. This result shows that Polygon-AI is useful.

4. Conclusions

In this technical paper, we introduced Polygon-AI, a data-based operation and equipment anomaly detection technology. By using this anomaly detection technology, it is possible to reduce the operation and equipment monitoring workload of operators and preempt operation and equipment problems before they occur. Therefore, aiming for operation and equipment innovations, we have developed and implemented an anomaly detection technology that makes it easy to explain and interpret the results and that can adapt to multiple operation modes.

Polygon-AI is currently being applied mainly to power generation and energy facilities. In the future, we intend to contribute more to our company's business by integrating Polygon-AI into our integrated platform for facility monitoring and by expanding the application of Polygon-AI to the steelmaking process as well.

References

1) Yairi, T.: Telemetry-Mining: A Machine Learning Approach to Anomaly Detection and Fault Diagnosis Problem for Space System. Journal of the Japanese Society for Artificial Intelligence. 21(1), 33–38 (2006)
2) Murakawa, M.: Anomaly Detection System Using Artificial Intelligence and Its Industrial Applications. Journal of the Atomic Energy Society of Japan. 59 (6), 335–339 (2017)
3) Shigemori, H.: Application of Data Science to Quality Control for Steel Products and Data Scientist Training. Systems, Control and Information. 62 (8), 311–316 (2018)

4) Natsumeda, M., Ochiai, K., Asakura, T., Hayashi, T.: Application of Invariant Analysis Technology to Large-Scale Physical System—Example of Application to Nuclear Power Plant Monitoring—. Journal of Digital Practice. 6 (3), 207–214 (2015)
5) Takahama, M., Mikami, N.: Detection of Abnormal Signs for Gas Turbine Power Plant. Journal of Quality Engineering Society. 20 (4), 45–51 (2012)
6) Lee, D. D., Seung, H. S.: Learning the parts of objects with nonnegative matrix factorization. Nature. (401), 788–791 (1999)
7) Sawada, H.: Nonnegative Matrix Factorization and Its Applications to Data/Signal Analysis. The Journal of the Institute of Electronics, Information and Communication Engineers. 95 (9), 829–833 (2012)
8) Nakao, K., Hayashi, T.: Spectrogram Feature Extraction and Anomaly Detection Using Non-negative Matrix Factorization. National Convention Record I.E.E. Japan. 2 (2019)

Masataka INOUE
Senior Manager
Control Development Section
Systems & Control Technology Dept.
Systems & Control Engineering Div.
Plant Engineering and Facility Management Center
20-1 Shintomi, Futtsu City, Chiba Pref. 293-8511