

Anomaly Detection in Predictive Maintenance using Dynamic Time Warping

Youngja Kim¹, Gyunghyun Choi²

¹ *Ph.D Candidate, Graduate School of Technology & Innovation Management, Hanyang University, Korea, puoupuou@hanyang.ac.kr*

² *Professor, Graduate School of Technology & Innovation Management, Hanyang University, Korea, ghchoi@hanyang.ac.kr*

Corresponding author: Gyunghyun Choi

Abstract: Manufacturing systems face the fundamental challenge of efficient operation by leveraging vast amounts of real-time data collected through technological advancements such as artificial intelligence and machine learning. Maintenance systems have evolved to predict and manage equipment failures in advance, with data-driven fault detection being a crucial technology. However, most related research has been limited to single equipment for specific processes, making the direct application in actual manufacturing settings that use various equipment models or types challenging. When using multifacility models, the most crucial aspect is the analysis of variations and errors in the data collected from each facility. To mitigate the risk associated with a sole vendor, different models of equipment is used strategically, even for the same functionality. Consequently, collecting temporally mismatched data is prevalent. The current methodology, which has been predominantly focused on a single-facility approach, faces limitations in its application when dealing with unstructured, unlabeled data, or temporally mismatched data obtained across multiple facilities. This study employed the dynamic time warping (DTW) method to analyze discrepancies in time-series data obtained from multiple equipment groups by leveraging similarity analysis of data peak matching for anomaly detection. Specifically, an approach called auto time windowing is adopted to extract signal periods based on the detailed signal analysis results of the process, enabling the application of DTW. The auto time windowing allows for the accurate automated analysis of signal period by overcoming the limitations of analysis errors caused by noise in the existing data using the threshold of the actual signal. This methodology is validated for two different equipment groups involved in a real-world production process, where parts are attached to products. The results of this study demonstrated an improvement over conventional time-series analysis methods such as the Euclidean method, addressing errors that may occur. This research enhances the analysis theory using DTW for the actual problem of data discrepancies among multiple equipment groups in the manufacturing field, which is not previously considered in existing predictive maintenance (PdM) theories. This validation through case studies effectively contributes to expanding the utilization of PdM.

Keywords: Dynamic Time Warping, Predictive Maintenance, Anomaly Detection, Multiple Facilities

1. Introduction

The manufacturing sector harnesses vast amounts of data through various digital devices and sensors to facilitate data-driven decision making and process improvement. These data are instrumental in

Received: September 12, 2023; 1st Review Result: October 13, 2023; 2nd Review Result: November 16, 2023
Accepted: December 26, 2023

improving product quality, reducing defects, cost savings, and extending the lifespan of products, thereby providing insights into performance management and decision making. Furthermore, they shorten the data processing time and improve predictive capabilities[1]. With the development of artificial intelligence, the importance of data control for automation and intelligent manufacturing using AI technology has increased. Through techniques such as deep learning and data mining, large-scale data systems are continuously being learned. Research is also being conducted on data-driven model correlations and algorithms[2]. Despite ongoing research into the capabilities and value of big data in manufacturing processes, challenges remain because of the complexity of processes, making it difficult to establish principled models and perform real-time data analysis for accurate predictions[3][4]. In particular, predictive analysis based on big data has been extensively researched from the perspectives of manufacturing operations, data comprehension, and analysis. However, further research is required to explore operational, performance, and cost-related aspects from a manufacturing standpoint[5].

Manufacturing production systems are complex, interconnected systems that encompass design, development, manufacturing, distribution, logistics, and so on. Within these systems, the manufacturing equipment area comprises various processes and facilities[6]. Despite the common goal of maximizing productivity, production and equipment maintenance schedules typically operate independently. Overemphasizing machine reliability or making frequent schedule changes can lead to confusion within the production system[7].

The foundation for the autonomous and collaborative control of equipment systems lies in maintenance systems. Maintenance involves a series of processes aimed at continuously preserving the original functionality of the equipment. The cost of maintenance accounts for 15%–60% of the total manufacturing cost. To improve efficiency and prevent system failures in advance, various studies have been conducted to monitor equipment conditions, detect failures in advance, and derive solutions [8–10]. The maintenance system has evolved from reactive maintenance, which addresses faults after they occur, to preventive maintenance (PM) and predictive maintenance (PdM), which focus on managing equipment before faults arise. Most of the relevant research has focused on individual equipment within specific processes, posing challenges for its direct applicability in real-world manufacturing settings characterized by the utilization of diverse equipment models or types. In actual manufacturing environments, multiple equipment groups are applied, necessitating realistic research to account for these scenarios[11][12].

This study conducted a comparative analysis to ascertain the applicability of research methodologies, which have previously focused on single-equipment models, in environments where multiple equipment groups operate with diverse models. On this basis, we proposed a method for analyzing data variances between equipment and validated its application through real-world cases involving multiple equipment groups. The multi-equipment analysis model proposed in this study will contribute to the enhancement of the utilization of data-driven equipment maintenance systems, addressing the shortcomings of PdM theory based on practical case analysis.

2. Literature Review

2.1 Failure Recognition and Prediction Research Trends

The core technologies in PM and PdM systems, which are part of equipment maintenance, are fault detection and prediction technologies aimed at preemptively identifying equipment failures. Regarding PM, maintenance schedules are established on the basis of the historical failure records of individual components, with intervals of 30 or 90 days. The component lifespan is determined on the basis of performance degradation due to aging[11][13]. However, PM involves replacing components on a predetermined schedule, even if they have not deteriorated or experienced failures, leading to ongoing

concerns about additional costs. Consequently, research is ongoing on modeling component replacement intervals, component costs, and equipment downtime management from a total cost perspective [14][15]. Furthermore, there is a need for efficiency models based on the condition of individual components, taking into account the actual deterioration status of components or optimization models for maintenance methods and execution times based on the failure type [16][17].

Conversely, PdM focuses on detecting failures before they occur to optimize costs. It relies on technologies such as the Internet of Things and big data for equipment condition monitoring, error diagnosis, data collection, and analysis. The main technology involves real-time monitoring based on data and the use of artificial intelligence and machine learning (ML) to optimize maintenance schedules. A key aspect for an effective maintenance schedule management is the clear analysis of remaining useful life. The ongoing research aims to propose a framework based on digital twins to enable real-time analysis of remaining useful life, and activities following the PdM pipeline have been suggested [18][19]. Specific components are equipped with sensors to secure data, and mathematical models for quantifying the deterioration status have been proposed. However, these studies are limited to single-equipment systems in specific processes, and research is required to address the challenges of multiple equipment groups in actual manufacturing environments and their expansion to other processes [20].

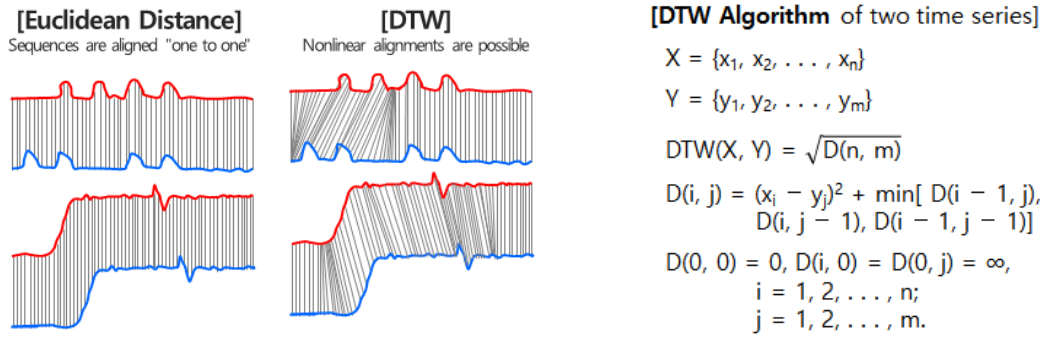
2.2 Variation Analysis among Multiple Facilities

When a product is designed and the design is finalized, the production process and equipment group required to create it are determined. At this point, a production operation plan is established, taking into consideration the production levels and specifications for each piece of equipment. When the production volume increases, multiple pieces of equipment are used. Even for the same equipment model group, there are differences in specifications because of continuous upgrades in performance and the addition of new features in newer models. For example, when purchasing equipment as a new model during equipment investment, both old and new models coexist. Some models add intelligent sensors to automatically generate data, and some models have not been fully deployed, resulting in differences in the collected data and functionality.

The collected data also mostly include irregular and unlabeled data. Data analysis is typically based on time series, and when dealing with inconsistent data, data analysis can become challenging [8][18][21]. Therefore, unstructured big data analysis is essential, and related research in this area is continuously growing [1].

Among various ML techniques, dynamic time warping (DTW) is extensively used for time-series clustering and classification. DTW compares the similarity between two signals through peak matching and allows the analysis of time-series signals through iterative distance measurements until the optimal match between the two signals is achieved [22-25]. Compared with the Euclidean method, which compares values at the same time points, as shown in [Fig. 1], DTW enables nonlinear mapping analysis between time-series signals. DTW is known for its simplicity and excellent real-time performance and is applied in various research fields such as speech recognition, pattern recognition, and error analysis. It is effective in applications such as network traffic analysis and anomaly detection, signal and fault type analysis in railway systems, bearing condition prediction, and fault recognition [23-26].

In the field of equipment maintenance, there is a need to analyze data that do not match over time series. In particular, in manufacturing environments in which multiple pieces of equipment are applied, it is necessary to extend the analysis method to analyze equipment state deviations and detect faults.

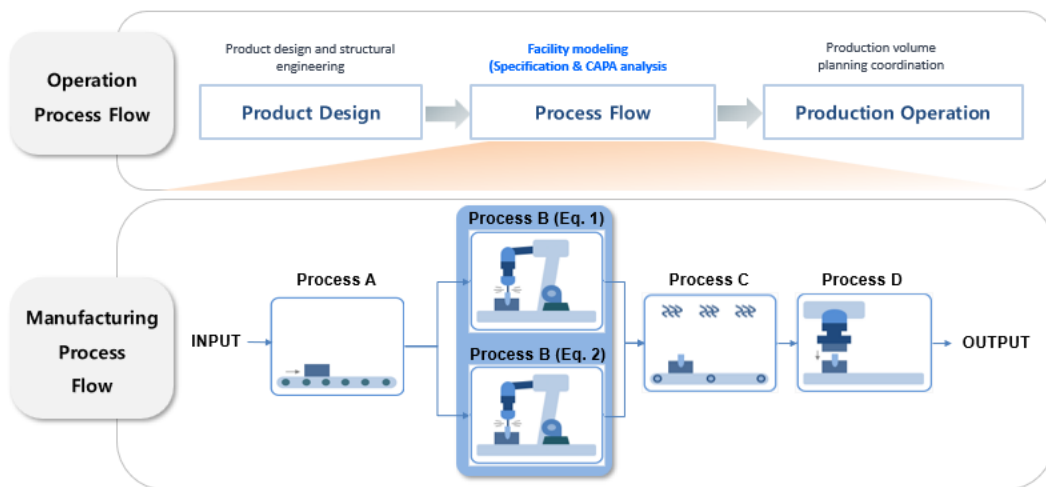


[Fig. 1] Distance Analysis between Euclidean and DTW

3. Research Methodology

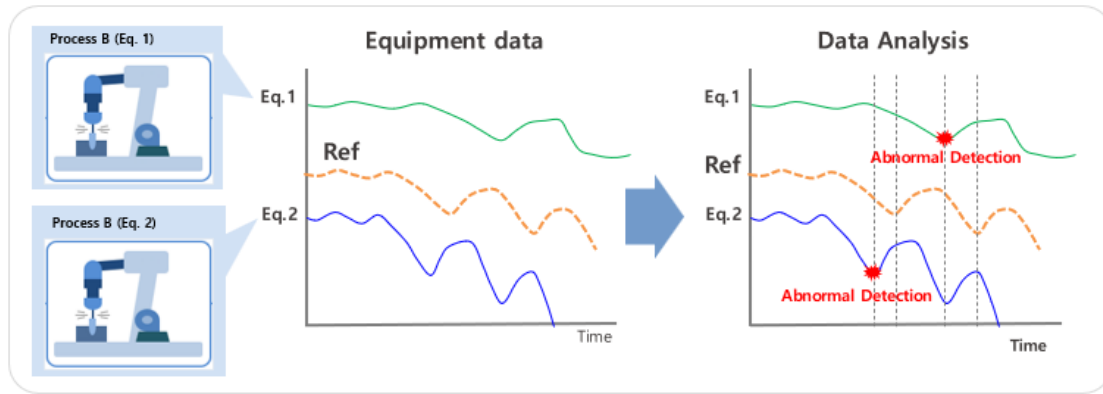
3.1 Dynamic Time Warping

The production process of customer products is depicted in [Fig. 2]. Initially, the product is designed, and its structure is planned. Subsequently, the necessary process sequence is determined, and an operational plan is developed on the basis of the production volume. Within this process, the process sequence is adjusted to align with the product specifications and the production capacity of each facility. Differences in the production capacity of each facility emerge, leading to the allocation of equipment models and the number of units required. In certain processes, such as Process B, multiple pieces of equipment are used to balance production quantities.



[Fig. 2] Process Flow of Manufacturing Operation

The data generated by each facility typically presents challenges in analysis because of time delays, temporal misalignment, irregularities, and the predominance of unstructured data with unspecified labels. In particular, in the case of multiple facilities, as illustrated in [Fig. 3], when data collected from each facility are temporally misaligned, false alarms can be triggered as anomalous events. PdM aims to detect equipment failures in advance by monitoring and collecting a vast amount of real-time data for analysis. However, dealing with temporally misaligned data poses constraints on precise time-point analysis and causal relationship interpretation, hindering its practical utility.



[Fig. 3] Data Analysis of Multiple Facilities

3.2 Results and Analysis

This study explores an approach for analyzing data patterns from multiple facilities using the DTW method, especially when dealing with temporally misaligned data. [Fig. 4] depicts the DTW algorithm, which calculates iterative distances between a reference signal and signals from each facility to derive optimal values. This logic allows for the assessment of the similarity between signals through peak matching. [Fig. 5] shows the results of pattern analysis for multifacility data using DTW, as depicted in (b), where peaks are matched to assess similarity. The data analysis results for the time series and DTW are presented separately in (c). When analyzed on the basis of the time series, Facility 1 exhibited a discrepancy score of 79 and Facility 2 showed a discrepancy score of 150. However, when DTW was applied, Facility 1 exhibited a similarity score of 88 and Facility 2 exhibited a similarity score of 89, confirming their similarity.

Algorithm of DTW (Fong, 2012)

```

DTW (R,E) {
// R and E are the time series with n and m time points(R=a1, b2, ...an, E=b1,b2, ...bn).
R is the reference data and E is the equipment data.
Similarity Matrix SM is the store of similarity measures such that SM[0,...n, 0,...m],
and p, q are loop index, diff is an integer.

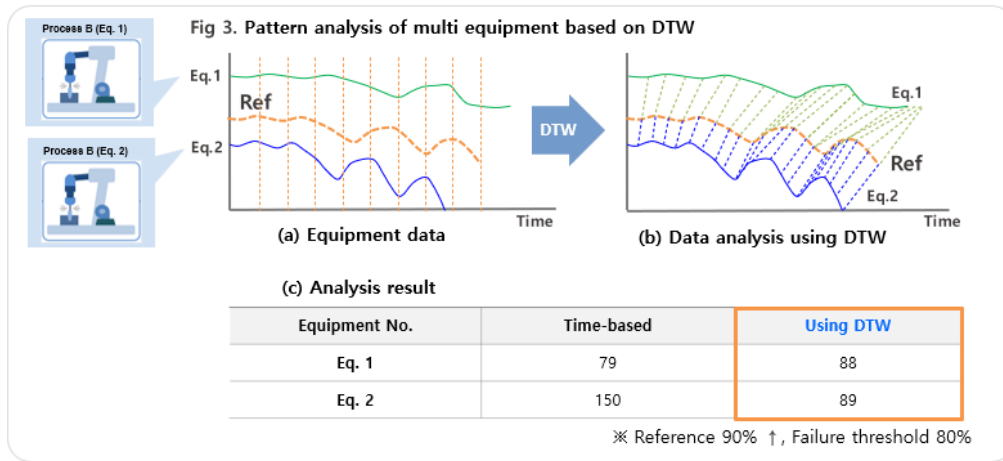
SM[0,0]=0
for p = 1 to m do
  SM[0,p] = ∞
for q = 1 to n do
  SM[q, 0] = ∞
  // Enter the differences between the two time-series stepwise into a similarity matrix
  using the pairwise method.

for q = 1 to n do
  for p = 1 to m to // A function that measures the distance between two points.
    diff = d(R[q], E[p]) // d is one of the distances of Minkowskui.
    SM[q,p] = diff + Min { SM[q-1, p]
                          SM[q, p-1]
                          SM[q-1, p-1]
    }

return SM[n, m]
}

```

[Fig. 4] Algorithm of DTW



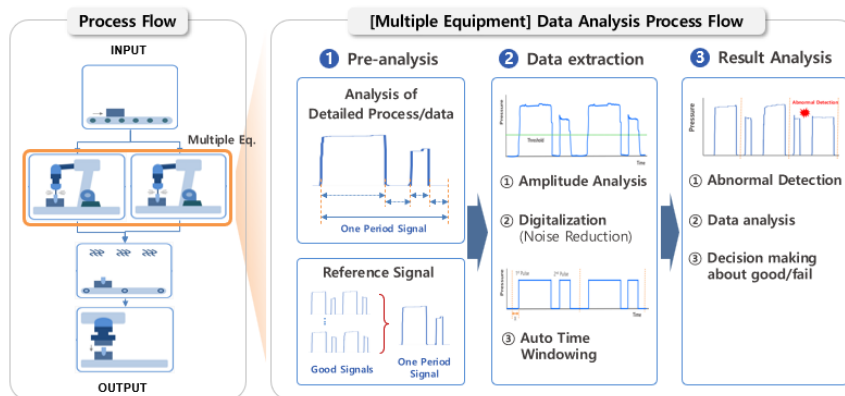
[Fig. 5] Pattern Analysis of Multiple Facilities based on DTW

4. Case Study

This study proposes a methodology for the analysis of multifacility data in actual manufacturing sites, which is scrutinized through a three-step procedure, as depicted in [Fig. 6]. Through such data analysis procedures, correcting data variances among multiple facilities is possible, which enable accurate analysis. Moreover, it is applicable not only to specific industries or processes, but can also be applied to analyze data collected from facilities with similar functionalities, without being constrained to a particular industry or process.

The first step involves the data analysis based on the detailed operations of the facilities and concurrently deriving a reference signal using good signals. The extraction of a one-period signal is crucial here. Subsequently, by comparing the reference signal obtained from the data of each facility, individual judgments are made. Therefore, the same methodology can be applied even with N facilities. Additionally, it represents a generalized approach applicable not only to the specific equipment studied in this case but also to other equipment and processes. The second step involves extracting data generated from the facilities and analyzing it using the auto time windowing technique proposed for noise removal. The final step includes progressing to the stage of recognizing outliers, analyzing their veracity, and making determinations.

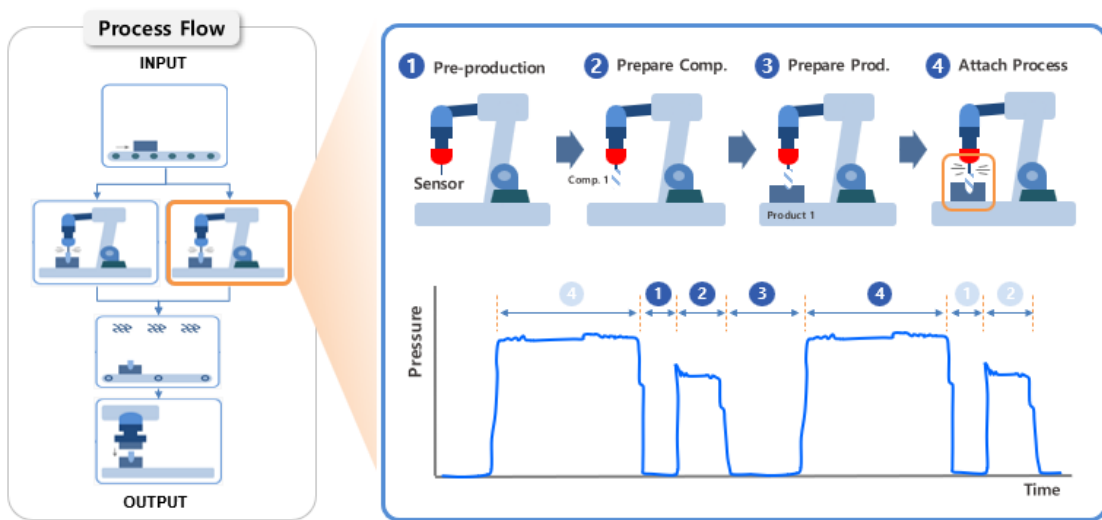
This case study has been validated using a three-step procedure, and as a future task, validating newly proposed methods, such as fast DTW, is planned to address the complexity and computational intensity of algorithm implementation arising from the feature extraction of data[27].



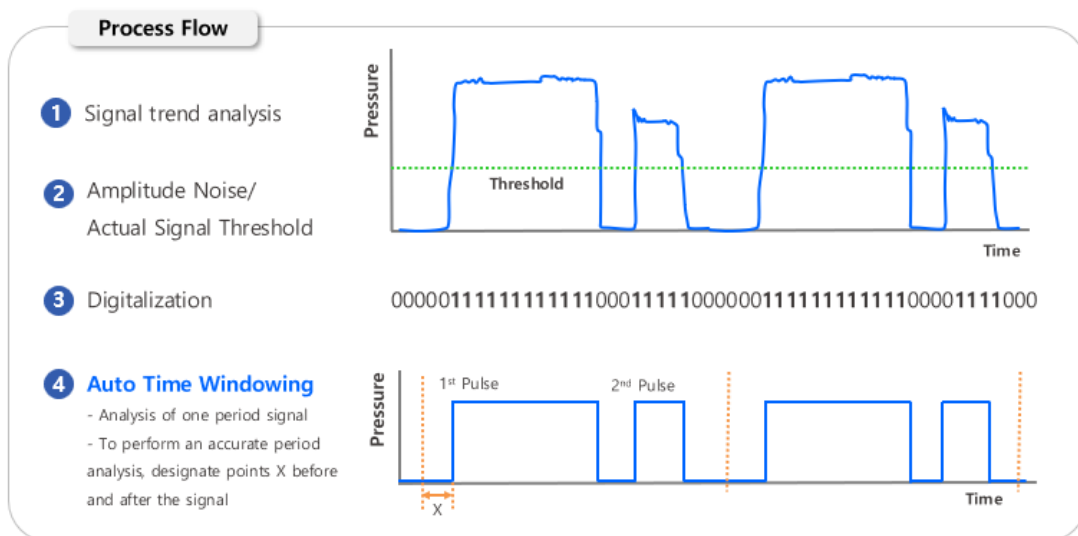
[Fig. 6] Process Flow for Data Analysis based on Multiple Facilities

4.1 Auto Time Windowing

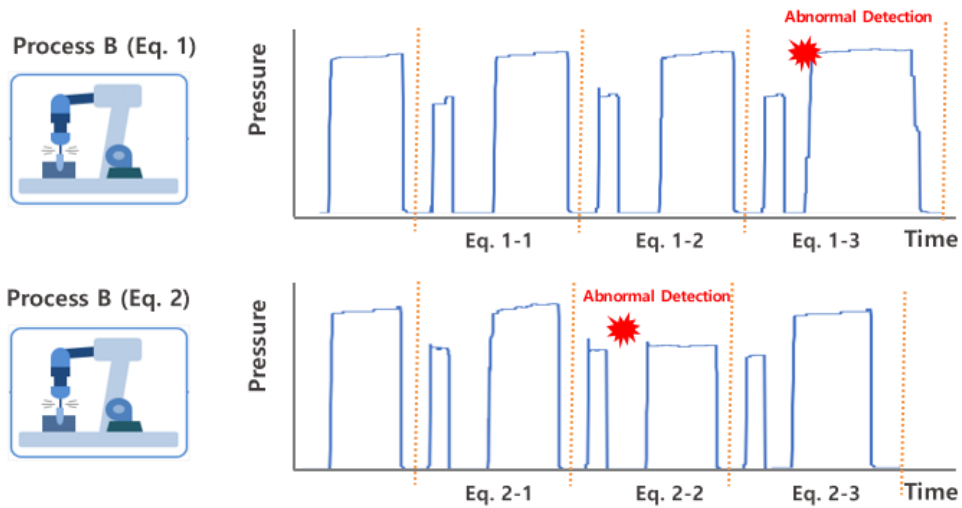
The following is a case analysis of the process of attaching components to a product, using both the Euclidean and DTW methods. Initially, data collected from the equipment undergo a preliminary analysis phase, as shown in [Fig. 7], to analyze detailed processes and data patterns within the operation. Data analyzed in this manner requires precise cycle analysis of signals; however, noise removal presents challenges. [Fig. 8] illustrates the automatic cycle analysis of signals through auto time windowing. Initially, the trend of the signal is analyzed, and the threshold values between actual signals and noise are determined. Subsequently, through digitalization, noise and signals are separated, and the cycles are analyzed. It is crucial to designate the points at which the signal changes, both before and after, as the X-points for a clear analysis of the starting and ending points of the cycle. This process is central to the analysis. [Fig. 9] shows the results of extracting the cycles of each dataset using auto time windowing.



[Fig. 7] Process Flow of Manufacturing Operation



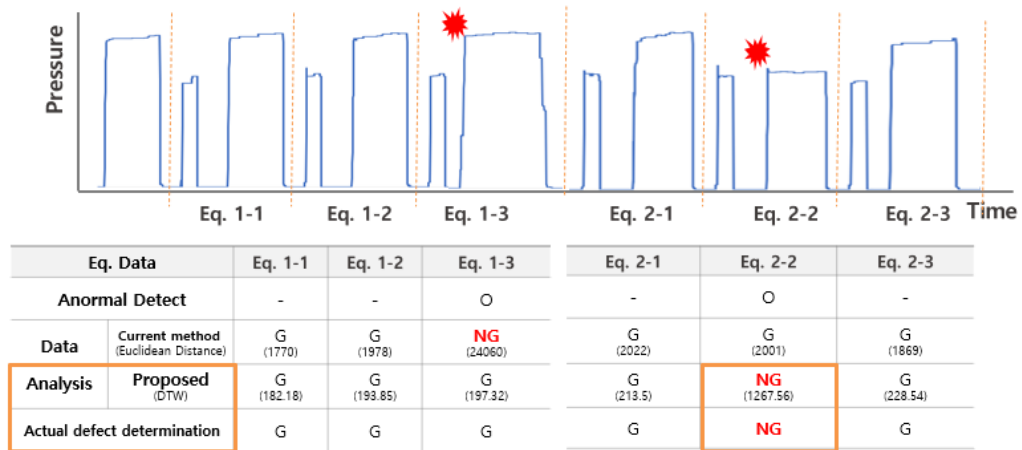
[Fig. 8] Auto Time Windowing



[Fig. 9] Period Analysis based on Auto Time Windowing

4.2 DTW Results and Analysis

The data analysis results of the Euclidean and DTW methods for data collected from Equipment Models 1 and 2 are shown in [Fig. 10]. Regarding Equipment 1, Signal 3 exhibited temporal misalignment, leading to previous Euclidean analysis classifying it as a defect. However, in reality, it was a good product unaffected by quality due to time delay. DTW analysis also classified it as "Good." Regarding Equipment 2, Signal 2 exhibited temporal alignment but had low pressure, resulting in adhesive defects. Euclidean analysis, based solely on temporal alignment, classified it as "Good." However, DTW analysis identified it as "No-Good."



[Fig. 10] Process Flow of Manufacturing Operation

5. Conclusions

The operation of equipment systems within a synergistic and interconnected manufacturing system is critical. In practical manufacturing systems, various equipment groups are operated to optimize

production capabilities and lead times. The equipment maintenance system has evolved into PdM, which is aimed at predicting failures in advance and is reliant on the fundamental technology of identifying failures based on collected data. However, unstructured data generated from individual equipment makes analysis challenging, particularly the efficient removal of temporally misaligned data or noise, which is a critical step in data analysis. This study uses the DTW methodology to analyze data discrepancies among multiple facilities for early anomaly detection through similarity analysis. Specifically, the process involves extracting data using DTW based on preliminary analysis of detailed process steps and signal analysis results, obtained through auto time windowing. Subsequently, the results were analyzed to determine the defect status. The methodology was validated through comparative verification of two equipment groups in the process of attaching components to products, confirming that DTW effectively identifies actual anomalies. This study presents a practical analysis model for multiple facilities in real-world manufacturing settings and, through case studies, supplements previously unaddressed real-world problems and theories within the PdM framework. These findings contribute to the enhancement of the utility of PdM practices. Especially in actual manufacturing settings, it can be deemed a valuable practical study because, strategically, multiple equipment models with the same functionality are employed to mitigate the risks associated with being a sole vendor. In subsequent research, the objective will be to authenticate recently suggested methodologies, including fast DTW, to tackle the intricacies and computational demands of algorithm implementation stemming from data feature extraction for precise fault detection and extensive dataset processing. These approaches focused on providing optimized algorithmic solutions through the streamlining of data of similar characteristics, thereby enabling real-time analysis of data generated in manufacturing processes.

References

- [1] N. Kumar, G. Kumar, R. K. Singh, Big Data Analytics Application for Sustainable Manufacturing Operations: Analysis of Strategic Factors, *Clean Technologies & Environmental Policy*, (2021), Vol.23, No.3, pp.965-989.
DOI: 10.1007/s10098-020-02008-5
- [2] J. Wang, Chuqiao Xu, J. Zhang, R. Zhong, Big Data Analytics for Intelligent Manufacturing Systems: A Review, *Journal of Manufacturing Systems*, (2022), Vol.62, pp.738-752.
DOI: 10.1016/j.jmsy.2021.03.005
- [3] A. Belhadi, K. Zkik, A. Cherrafi, S. M. Yusof, S. El fezazi. Understanding Big Data Analytics for Manufacturing Processes: Insights from Literature Review and Multiple Case Studies, *Computers & Industrial Engineering*, (2019), Vol.137, 106099.
DOI: 10.1016/j.cie.2019.106099
- [4] P. B. Radu, E. Bobolea, F. Nordbjerg, T. Baattrup-Andersen, N. Iftikhar, Data Analytics for Smart Manufacturing: A Case Study, *Proceedings of the 8th International Conference on Data Science Technology & Applications*, (2019)
DOI: 10.5220/0008116203920399
- [5] R. Dubey, A. Gunasekaran, S. J. Childe, C. Blome, T. Papadopoulos, Big Data and Predictive Analytics and Manufacturing Performance: Integrating Institutional Theory, Resource-Based View and Big Data Culture, *British Journal of Management*, (2019), Vol.30, No.2, pp.341-361.
DOI: 10.1111/1467-8551.12355
- [6] S. Geng, X. Wang, Predictive Maintenance Scheduling for Multiple Power Equipment Based on Data-Driven Fault Prediction, *Computers & Industrial Engineering*, (2022), Vol.164, 107898.
DOI: 10.1016/j.cie.2021.107898
- [7] S. Wang, J. Yu, An Effective Heuristic for Flexible Job-Shop Scheduling Problem with Maintenance Activities, *Computers & Industrial Engineering*, (2010), Vol.59, No.3, pp.436-447.
DOI: 10.1016/j.cie.2010.05.016

- [8] A. Bakdi, N. B. Kristensen, M. Stakkeland, Multiple Instance Learning with Random Forest for Event Logs Analysis and Predictive Maintenance in Ship Electric Propulsion System, *IEEE Transactions on Industrial Informatics*, (2022), Vol.18, No.11, pp.7718-7728.
DOI: 10.1109/TII.2022.3144177
- [9] S. Arena, E. Florian, I. Zennaro, P. F. Orrù, F. Sgarbossa, A Novel Decision Support System for Managing Predictive Maintenance Strategies Based on Machine Learning Approaches, *Safety Science*, (2022), Vol.146, 105529.
DOI: 10.1016/j.ssci.2021.105529
- [10] T. Zonta, C. A. da Costa, R. da Rosa Righi, M. J. de Lima, E. S. da Trindade, G. P. Li, Predictive Maintenance in the Industry 4.0: A Systematic Literature Review, *Computers & Industrial Engineering*, (2020), Vol.150, 106889.
DOI: 10.1016/j.cie.2020.106889
- [11] J. Zhao, C. Gao, T. Tang, A Review of Sustainable Maintenance Strategies for Single Component and Multicomponent Equipment, *Sustainability*, (2022), Vol.14, No.5, 2992.
DOI: 10.3390/su14052992
- [12] T. Nemeth, Prima-X: A Reference Model for Realizing Prescriptive Maintenance and Assessing Its Maturity Enhanced by Machine Learning, *Procedia CIRP*, (2018), Vol.72, pp.1039-1044.
- [13] J. Zenisek, F. Holzinger, M. Affenzeller, Machine Learning Based Concept Drift Detection for Predictive Maintenance, *Computers & Industrial Engineering*, (2019), Vol.137, 106031.
DOI: 10.1016/j.cie.2019.106031
- [14] B. Lu, X. Zhou, Opportunistic Preventive Maintenance Scheduling for Serial-Parallel Multistage Manufacturing Systems with Multiple Streams of Deterioration, *Reliability Engineering & System Safety*, (2017), Vol.168, pp.116-127.
DOI: 10.1016/j.ress.2017.05.017
- [15] J. Gan, W. Zhang, S. Wang, X. Zhang, Joint Decision of Condition-Based Opportunistic Maintenance and Scheduling for Multi-component Production Systems, *International Journal of Production Research*, (2022), Vol.60, No.17, pp.5155-5175.
DOI: 10.1080/00207543.2021.1951447
- [16] H. Jafar-Zanjani, M. Zandieh, M. Sharifi, Robust and Resilient Joint Periodic Maintenance Planning and Scheduling in a Multi-factory Network under Uncertainty: A Case Study, *Reliability Engineering & System Safety*, (2022), Vol.217, 108113.
DOI: 10.1016/j.ress.2021.108113
- [17] J. Liu, D. Yu, Y. Hu, H. Yu, W. He, L. Zhang, Digital Twin-Driven Multi-objective Optimization Production Scheduling with Restraint Tool Resources, 2021 7th International Conference on Computer and Communications (ICCC), (2021)
DOI: 10.1109/ICCC54389.2021.9674414
- [18] M. Pech, J. Vrchota, J. Bednář, Predictive Maintenance and Intelligent Sensors in Smart Factory: Review, *Sensors*, (2021), Vol.21, No.4, 1470.
DOI: 10.3390/s21041470
- [19] S. Sharanya, R. Venkataraman, G. Murali, Edge Ai: From the Perspective of Predictive Maintenance, In *Applied Edge Ai*, (2022)
- [20] M. Sharifi, S. Taghipour, Optimal Production and Maintenance Scheduling for a Degrading Multi-failure Modes Single-Machine Production Environment, *Applied Soft Computing*, (2021), Vol.106, 107312.
DOI: 10.1016/j.asoc.2021.107312
- [21] E. Mohebi, Identifying the Substantial Big Data (Bd) Aspects to Improve, 5th international conference on industrial engineering and operations management, (2020)
- [22] H. Kim, S. Park, S. Ki, Time-Series Clustering and Forecasting Household Electricity Demand Using Smart Meter Data, *Energy Reports*, (2023), Vol.9, pp.4111-4121.
DOI: 10.1016/j.egyr.2023.03.042
- [23] J. Guo, Zetian Si, Y. Liu, Jiahao Li, Y. Li, J. Xiang, Dynamic Time Warping Using Graph Similarity Guided Symplectic Geometry Mode Decomposition to Detect Bearing Faults, *Reliability Engineering & System Safety*, (2022), Vol. 224,

108533.

DOI: 10.1016/j.ress.2022.108533

- [24] T. Han, X. Liu, A. C. C. Tan, Fault Diagnosis of Rolling Element Bearings Based on Multiscale Dynamic Time Warping, *Measurement*, (2017), Vol.95, pp.355-366.
DOI: 10.1016/j.measurement.2016.10.038
- [25] S. Huang, F. Zhang, R. Yu, W. Chen, F. Hu, D. Dong, Turnout Fault Diagnosis through Dynamic Time Warping and Signal Normalization, *Journal of Advanced Transportation*, (2017), Vol.2017, pp.1-8.
DOI: 10.1155/2017/3192967
- [26] B. M. A. Diab, H. Binsalleh, S. Lambotharan, K. G. Kyriakopoulos, I. Ghafir. Anomaly Detection Using Dynamic Time Warping, 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), (2019)
- [27] Nien-Che Yang, and Jen-Ming Yang. Fault Classification in Distribution Systems Using Deep Learning With Data Preprocessing Methods Based on Fast Dynamic Time Warping and Short-Time Fourier Transforms, *IEEE Access*, (2023), Vol.11.
DOI: 10.1109/ACCESS.2023.3288852