Matrix profile implementation perspective in Industrial Internet of Things production maintenance application

J Pizoń, M Kulisz and J Lipski
Lublin University of Technology, Management Faculty, ul. Nadbystrzycka 38, 20-618 Lublin, Poland
j.pizon@pollub.pl

Abstract. The matrix profile processing is considered for the implementation of production maintenance tasks in the context of data acquisition by industrial Internet of Things solutions. The prospective implementation of the matrix profile data structure is verified through a dedicated case study, presenting the method of processing non-labelled data registered by sensors. The case study demonstrates the functionality of the profile and indicates the prospects of their applications in the field of production maintenance.

1. Introduction
Computer systems in production eliminate problems and errors. In a production company, IT systems support planning, control, and maintenance while being implemented at almost every level of the organisation, from the tactical level to the management level. Thanks to their modular structure, IT systems provide the flow of information inside the organisation and outside – to its partners. In operation, they use and implement computational (mathematical, IT) or organisational (management) methods [1]. The methods rely on the availability of data relating to the manufacturing system and their task is to process them in order to make decisions adequate to the dynamically changing state and status of the manufacturing system (Figure 1). The state of the system is assumed to provide diagnostic information regarding whether the system works properly (considering the expected behaviour) or not [2], whereas, the status determines whether a given production system is working. Here, the problem that arises is that the answers to the questions must be supplied at the right time and place while being specific, i.e. concerning a single machine as well as the entire system. To provide the management function to managers, drill-down data are necessary in order to give insight into an organisation from every point of view. However, uncontextualised data do not constitute a business case. Data is critical but effective data processing is crucial to retrieving management information. Therefore, an information system develops both in terms of data aggregation and data processing [3]. Valuable management information is obtained by referring to a current problem requiring a decision. Management decisions need to correspond to the changing requirements of the environment and the internal state of the production system. The basis for the reaction is feedback.

Owing to this mechanism, the managers can evaluate the consequences of the decision taken. Feedback gives a possibility to measure the change and to verify its results. If the changes did not bring the desired results, they should be abandoned, otherwise, they should be continued. The better and more complete information about the status and state of the production system, the better decisions can be made. Computer systems are thus applied to support the process of data acquisition. This opens the door
to many applications in the monitoring and control of manufacturing and is also the basis for the development of modern IT support systems for production systems – especially in the area of efficiency. These technologies include:

- Internet of Things (IoT),
- Big Data,
- Cyber-Physical Systems (CPS),
- Cloud Computing (CC),
- Fog Computing (FG),
- Digital Twin (DT),
- Industrial Artificial Intelligence (IAI).

Figure 1. Manufacturing system supported by management support system based on IIoT.

Some of these technologies focus on data aggregation and transfer. Others focus on providing computing power for processing or creating a model that uses the acquired data to map objects in digital space and equipping it with a certain level of intelligence (Table 1).

| Aggregation and transfer          | Computation power          | Modelling and intelligence                                      |
|----------------------------------|----------------------------|-----------------------------------------------------------------|
| Internet of Things (IoT)         | Cloud Computing (CC)       | Industrial artificial intelligence, IAI                         |
| Big Data                         | Fog Computing (FG)         | Digital twin (DT)                                               |
|                                  |                            | Cyber-Physical Systems (CPS)                                    |

Table 1. Information technology categories. Source: own work.

The application of these modern technologies on the basis of production systems is referred to as the fourth industrial revolution or, in short, industry 4.0. The development of industry 4.0 systems is based mainly on IT solutions dedicated to industrial applications. Industrial IT is an implementation of native IT technologies suitable for the requirements of the industrial activity domain. These technologies provide an interdisciplinary combination of manufacturing engineering, information systems, mechatronics, communication technologies and data analytics.

Industry 4.0 solution is a set of IT functionalities, selected and connected to the reported business needs of a production company [4]. The term Industry 4.0 was created as a strategic initiative of the German government, which was adopted as part of the "Action Plan for an Advanced Technology Strategy 2020" in 2011. The discussion on Industry 4.0 began in Germany and spread to other countries, taking various names, e.g. industry 4.0 in Europe, smart manufacturing in the USA, or smart factory in Asia. The essence of the term industry 4.0 is to indicate that today the fourth industrial revolution is taking place [5]. The revolution is to lead to the creation of new-generation factories using information...
technologies for comprehensive multidimensional support of industrial processes [6]. Nevertheless, to make it possible, it is necessary to use advanced algorithmic methods to process large sets of non-labelled data. Both methods and tools are needed to identify adverse events such as disruptions and anomalies that may herald upcoming downtime and breakdowns. The article considers through a case study whether the features and functionalities of the matrix profile can be the basis for the application of maintenance strategies.

2. Literature background

2.1. Industrial Internet of Things

The revolution is possible mainly due to access to data. The solution that has significantly changed the way in which data on production facilities is acquired today is commonly referred to as the Industrial Internet of Things. The IIoT focuses on production processes, asset management, maintenance, monitoring and control. The role of this technology is to aggregate data on means of production in real time [7]. The data obtained in this way is the basis for the implementation of maintenance activities (failure prediction) and verification of the effectiveness of the production system [8]. The IIoT provides the basis for providing digital feedback for manufacturing management processes. IIoT enables monitoring of the production system even in those areas where there no digital control or control systems are present: traditional machine tools, means of transport or production in progress [9].

Data acquisition from the production process can be done manually or automatically with the use of embedded machine systems [10]. The selected harvesting method depends on the size and complexity of the manufacturing system being managed. Unless processed, the signals captured by the sensors are useless for higher-level applications. The sensors allow for the acquisition of various signals (power, torque, temperature, vibration or acoustic emission) that correlate with the state and status of manufacturing processes. Which of these signals is critical for the implemented process of the production system, and more precisely for a given technological process carried out by a given machine, results from the technology or an adopted maintenance strategy.

From the processing perspective, the signals are represented as records – time series. A time series is a sequence of observations showing the development of a phenomenon under study in subsequent periods (days, months, quarters, years, etc). Time series analysis has two main goals. The first is to discover the nature of the phenomenon represented by the sequence of observations. The second is forecasting (predicting future values of the time series). These goals require identifying and describing the elements of the time series. Once established, the pattern can be applied to other data, that is, used in the theory of the studied phenomenon, for example, seasonal commodity prices [11]. Therefore, only the application of the processing algorithm combines the obtained data and production events. The selection of the algorithm requires a simultaneous understanding of the monitored process and the characteristics of the processed signal [12].

2.2. Matrix Profile

Given the above requirements, particular attention is paid to the use of the distance method to calculate the characteristics of the data stream known as the matrix profile (MP). The matrix profile is the distance vector between each time series sequence and its closest neighbour. The vector and its calculation algorithms were proposed in 2016 by a team led by Eamonn Keogh from the University of California, Riverside and Abdullah Mueen from the University of New Mexico [13]. Works on the matrix profile respond to the need for fast and accurate data processing of time series associated with the increase in the number of applications and popularity of the Internet of Things and cloud computing technologies. Therefore, the matrix profile as a promising way of using the distance method in the scope of the dissertation issues will be characterised in detail and used in this work.

The algorithms for calculating the characteristics of the time series developed by simplifying the mathematical operations of the distance method significantly reduced the complexity of calculations and enabled calculating the profile much faster than in a naive manner – as referred by the vector name. This
name resulted from the fact that one of the least effective ways of calculating the distance vector of two-time series sub-sequences would be to compute the full distance matrix of all sequences in one-time series with all the sequences in another time series and to extract the smallest value for each element of a given sequence. This, of course, is associated with high complexity and therefore long computation times for all distances. Both the matrix profile and the profile index allow exploring the properties of a given time series (Figure 2). As indicated, it can be used to show patterns (motifs) as well as disturbances (anomalies). With regard to the matrix profile, the pattern is a repeated sequence in a time series, and the disturbance is a sequence that does not repeat, as shown in Figure 2.

The matrix profile calculates distances in the Euclidean space, which means that a distance close to 0 is most similar to another sequence in the time series, and a distance far from 0 is not similar to any other sequence. Extracting the shortest distances shows the patterns, while the greatest distances show discords.

The features of the matrix profile include [13]:
- Accuracy – in the case of discovering motifs, detecting discrepancies or series of time patterns – profile-based methods do not give false results or false references,
- Limited parameterisation – compared to other methods, generating a profile does not require detailed parameterisation, it is enough to indicate the comparison window m,
- Does not require the provision of linear memory space for calculations appropriate to the size of the set – only the matrix profile is counted, which requires much less space than processing the entire sample, thereby enabling processing massive data sets in the main memory,
- Can be created with the help of various algorithms,
- Can be updated – after calculating the matrix profile for a data set, it can be gradually updated,
- No need to define the similarity threshold – does not require the user to set similarity/distance thresholds,
- Parallelism – the matrix can be generated on both multi-core processors and distributed systems,
- The profile can be calculated in deterministic time – considering only the length of the time series, you can accurately predict in advance how long it will take to calculate the matrix profile.

The key advantages of the matrix profile vector and its computation method are that it is domain-independent data that provide an accurate solution. Only one parameter of the length of the sequence range window against which the comparison will be made is required. This opens up great prospects for solutions in the field of time series data analysis. What is more, the profile itself can be used in the feature engineering process. That can later become the base for artificial intelligence solutions and digital twin solutions implementation.

![Figure 2. Characteristics of the structure of the distance matrix profile [14].](image-url)
2.3. Production maintenance

The effectiveness of production systems depends on the adopted strategy and the maintenance strategy. Maintenance involves a series of repair and monitoring activities that include condition analysis, routine check-up, inspection, and repair [15], by primarily the maintenance department as well as all employees. The activities in the field of machine maintenance relate to the prevention of losses related to unplanned downtime caused by a breakdown or other event preventing scheduled work. In the event of unplanned downtime, the costs are incurred in two dimensions, both in terms of the cost of repair and the cost of delays in production. The cost of implementing the production will correspond to the delay. If the damaged spare element is in the spare parts warehouse – the repair will be carried out immediately. Otherwise, if it is not available and needs to be imported, the total repair time will be significantly longer. Taking the machine out of the production cycle reduces the possibilities of fulfilling current and future orders, which comes at the cost of lost opportunities. Counteracting such situations involves taking coordinated actions as part of the maintenance strategy. Thanks to the implementation of the strategy, it is possible to reduce the costs of failure and service as well as downtime, while extending the efficient operation time. The strategies include various activities such as: failure prediction, failure diagnosis, failure detection, failure type classification, a recommendation of corrective actions, and maintenance after failure. The maintenance strategy includes [16]:

- reactive maintenance – slowdowns and failures are removed when they occur,
- preventive maintenance – equipment maintenance is carried out in accordance with a pre-planned schedule in order to replace or maintain parts prior to failure,
- predictive maintenance – the operation of a device is monitored with sensors, and historical data is analysed to predict the pending failure and solve the problem proactively,
- prescriptive maintenance – the equipment performance is constantly monitored by sensors, and the data is subsequently analysed by advanced software to propose specific maintenance activities that ensure optimal equipment uptime. Specifically, prescriptive maintenance includes expert knowledge modelling, machine learning, predictive data analysis, and semantic reasoning to streamline and automate decision-making processes. Optimal selection and proposing the right strategies, tactics and action plans to anticipate and solve problems concern a given manufacturing system and individual machines. Prescriptive maintenance uses descriptive, diagnostic, and predictive analyses to understand past events to predict the likelihood of future events and the potential impact of each maintenance decision on the manufacturing system and related business processes.

Maintenance strategies are classified by criteria of implementation complexity and the expected value for the enterprise. Deciding on an optimal strategy for a given company depends on the specific issue to tackle. The form of the response is determined by the result of adopting a given strategy. The adopted strategy will recommend a set of actions aimed at providing information, conducting an analysis or recommending actions [17].

3. Case study

The indicated strategies focus on introducing measures to prevent failures and their negative effects. The use of new information technologies opens up a new possibility for managing production systems. The availability of big data in itself does not guarantee that every issue will be noticed and captured. This particularly concerns symptoms of impending failures signalled by repeated deviations from the standard indications, which may in fact sometimes represent the standard scope.

Developments in IT lead to the emergence of new data and the need for their processing. Hence, this paper attempts to assess the prospect of implementing matrix profiles as part of maintenance system solutions, based on data obtained within the industrial Internet of Things, focusing on the functionality of identifying anomalies. The used dataset is composed of untagged data – analysed using a matrix profile.
3.1. Data source
The case study used a dataset provided by the Prognostics CoE at NASA Ames with use of C-MAPSS simulation engine [18]. The dataset contains records of several sensor channels to characterise faulty fan revolution in a gas turbine engine (turbofan engine). The data is divided into training and test subsets consisting of multiple non-labelled time series – that within this article will be treated as IIoT data gathered in real time. It is assumed that the sensors transmit real-time data that are accumulated and processed.

The dataset presents data acquired from 100 different gas turbine engines of the same type. Readings of each engine start with different degrees of initial wear and manufacturing variation, which is unknown. The engine operates normally at the start of each time series and develops a fault at some point during the series. In the training set, the fault grows in magnitude until system failure. Time is measured in cycles. Cycles indicate the time between each timeseries’ record. Moreover, the selected dataset contains a vector of Remaining Useful Life (RUL) provided in cycle values of each engine. The engine timeseries sensor data and the RUL open the prospect of validating the method by comparing RUL with the matrix profile.

3.2. Research results
In order to conduct the research, an exploratory analysis was performed. Each sensor measure was evaluated. From the case study perspective, evaluating the method makes sense in relation to at least one measure. For this reason, to run processing with matrix profile temperature (in °R) sensor data was used. The temperature was measured at the outlet of the low-pressure compressor chamber. The exploratory analysis showed that with the passage of successive cycles, the range of recorded records continued to increase – as did its average (Figure 3). Similar characteristics were observed for the remaining engines. Thus, it can be assumed that its wear and tear increased along with the increase in the number of cycles.

![Figure 3. Gas turbine engine number 16 sensor measurements.](image)

![Figure 4. Matrix profile for gas turbine engine number 16 measurements.](image)
Next, for each engine for a given metric, the matrix profile was calculated. To determine the analysis window required by the matrix profile, refractoriness under load (RUL) values were used. It was assumed for the case study that the analysis window would be the average of the RUL results, that is 75 cycles. Considering given assumptions a profile calculation was performed with the use of the tsmp library implemented in the R language. Calculating the profile (Figure 4) for the following engines took up to about 0.1 seconds per engine data. In the case of engine 16, the calculated matrix profile indicates for cycles presented as index 43 and 130 approaching zero indicating that there is a repeating pattern in the data.

Subsequently, to identify emerging anomalies, the profiles were verified in terms of the occurrence of discords (non-recurring sequences in time series). The cycle index in which it occurred was recorded for each disorder. The algorithm for searching for disorders always indicates the sub-sequence that to the greatest extent mismatches the studied sub-sequence (Figure 5), which is the basis for labelling it as an anomaly. The number of disturbance cycles was compared with the expected times of remaining use (Figure 6).

![Discord Discover](image1)

Figure 5. Discord for gas turbine engine number 16 measurements.

![Chart](image2)

Figure 6. Comparison of RUL and Anomaly.

Identification of anomalies on the basis of the disturbance detection using the matrix profile calculations indicated that in 78% of cases the strongest disturbance occurred before the estimated RUL.
In other cases, in the analysed data, the moment of failure was diagnosed after the estimated time of RUL. Such a result is related to the fact that the process of engine wear is a multidimensional problem – dependent on a greater number of factors. In the analysed set, anomalies occurred on average within the 39th cycle, which was clearly related to the start of work by a given engine.

The analysis describes the operation of the machine until failure. The results of the cycle with the last disturbance in relation to the predicted RUL were combined, which enabled calculating the cycle difference between the last disturbance and the sample length. The obtained result, i.e. the number of cycles from the last disturbance to failure was compared with the RUL. The results of the analysis are presented in Figure 7.

![Figure 7. Comparison of RUL and Anomaly.](image)

The combination of these two approaches indicated that the profile and the RUL successfully predicted the moment of failure with an average accuracy of 15 cycles, while the median was 9 cycles after failure, respectively.

3.3. Discussion of results

In the presented case study, due to the assumptions made regarding the verification of the calculation method, the focus was on only one metric. Nevertheless, analysing the available range of data, indications of disorders were obtained in accordance with the suspected characteristics.

On the other hand, given the assumptions and the lack of domain knowledge on engine operation, the adopted analysis window may have been too large in relation to the analysed data. Moreover, this window has been established for all engines disregarding the settings for the following cycles. Apart from the individual parameters of the engines. These aspects definitely influenced the fact that the results of the method application differ significantly from the indications of the RUL cycles. The number of cycles between the occurrence of the strongest disturbance or the last disturbance is quite large. Due to this, in industrial applications, such results cannot be applied – because of inadequate accuracy.

However, it should be noted that the latest sequence of disorder was indicative of the cycle that began the process of change in the average amplitude – leading to failure. This fact discovered in the data may already be the basis for the application and the use of an appropriate maintenance strategy. In the event of detecting such a moment, maintenance services should take steps to verify the indications of a given sensor.

Moreover, the possibility of discovering disturbances in the data makes it possible to identify anomalies in the obtained disturbance patterns. There are possible applications of monitoring systems which, after detecting a similar record, will alert the maintenance services. It can be suspected that collecting data on disturbances will enable the development of a prevention system – identifying anomalies that indicate failures before they occur. Therefore, when using a profile, it is necessary to
carry out calculations and to define a specific strategy of its application. Moreover, high-quality data should also be obtained along with a description of their full characteristics. It is then that one can expect the perspective of the profile application to be even greater.

What is more, from the perspective of computation complexity, the case study has shown that for the data series consisting of over 300 cycles the calculations lasted up to 1 second on a typical computer. That opens the possibility for analysing much larger data sets in a satisfactory time in the future.

So far, researches on the profile have focused on expanding its functionality and computational description [19]. At the same time, some application articles have already been published. They use the profile to search for anomalies within power systems [20] or to analyse seismographic data in earthquake detection [21]. The number of applications is still small so there is a new large application perspective both for research and practical industrial application.

4. Conclusions
In conclusion, the prospect of matrix profile implementation on the basis of maintenance systems is quite promising. In terms of applicability, the functionality of indicating anomalies draws special attention. This feature of matrix profile can be used to deliver a reactive maintenance strategy that focuses on implementation rules, procedures, but also automatic mechanisms that will directly indicate the occurrence of failure. In addition, it can be the basis for the implementation of advanced strategies (preventive, predictive and prescriptive maintenance) taking into account the fact that using the functionalities of searching for patterns and anomalies as training data. It is possible to develop artificial intelligence solutions that will enable predicting upcoming changes. The scope of applications seems unlimited; however, it should be remembered that the basis for these applications is high-quality data obtained in the production environment.

The matrix profile is also a computational technique. In this respect, the matrix profile is the answer to the problem of processing large non-labelled datasets, such as time series of physical parameters of the processes being carried out. Taking into account the features and functionalities of the profiles, it is possible to process data from many sources (sensors or systems) within a specified acceptable time. Thus, the implementation perspective, both in terms of the complexity of calculations and results, corresponds to the needs of processing production data.

Taking into account the obtained results, further research will focus on using the matrix profile to develop a predictive system – allowing the identification of anomalies that indicate failures prior to their occurrence.

References
[1] Zawadzka L, Badurek J and Łopatowska J 2012 Inteligentne systemy produkcyjne Algorytmy koncepcje zastosowań (Gdańsk: Wydawnictwo Politechniki Gdańskiej)
[2] Lipski J 2013 Diagnostyka procesów wytwarzania (Lublin: Wydawnictwo Politechniki Lubelskiej)
[3] Raptis T P, Passarella A and Conti M 2019 Data Management in Industry 4.0: State of the Art and Open Challenges IEEE Access 7 97052-97093 https://doi.org/10.1109/ACCESS.2019.2929296
[4] Zhong R Y, Xu X, Klotz E and Newman S T 2017 Intelligent Manufacturing in the Context of Industry 4.0: A Review Engineering 3(5) 616-630 https://doi.org/10.1101/J.ENG.2017.05.015.
[5] Weyer S, Schmitt M, Ohmer M and Gorecky D 2015 Towards Industry 4.0 Standardization as the crucial challenge for highly modular, multi-vendor production systems IFAC-PapersOnLine 48(3) 579–584
[6] Roblek V, Meško M and Krapčič A 2016 A Complex View of Industry 4.0 (SAGE Open) https://doi.org/10.1177/2158244016653987
[7] Pizoń J, Kłosowski G and Lipski J 2019 Key role and potential of Industrial Internet of Things (IIoT) in modern production monitoring applications MATEC Web of Conferences 252 09003 https://doi.org/10.1051/matecconf/201925209003
[8] Sisinni E, Saifullah A, Han S, Jennehag U and Gidlund M 2018 Industrial Internet of Things: Challenges, Opportunities, and Directions *IEEE Transactions on Industrial Informatics* **14**(11) 4724-4734 https://doi.org/10.1109/TII.2018.2852491

[9] Daji D, Ghule K, Gagdani S, Butala A, Talele P and Kamat H 2020 Cloud-Based Asset Monitoring and Predictive Maintenance in an Industrial IoT System *2020 International Conference for Emerging Technology (INCET)* (Belgaum, India) pp 1-5 https://doi.org/10.1109/INCET49848.2020.9154148

[10] Beňo L, Pribiš R and Leskovský R 2019 Processing data from OPC UA server by using Edge and Cloud computing *IFAC-PapersOnLine* **52**(27) 240-245 https://doi.org/10.1016/j.ifacol.2019.12.645

[11] Analiza szeregów czasowych StatSoft Electronic Statistic Textbook https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fsttimser.html [Access date: 09.05.2020]

[12] Cao W, Jiang P, Lu P, Liu B and Jiang K 2017 Real-time data-driven monitoring in job-shop floor based on radio frequency identification *International Journal of Advanced Manufacturing Technology* **92**(5–8) 2099–2120

[13] *Introduction to Matrix Profiles - Towards Data Science* https://towardsdatascience.com/introduction-to-matrix-profiles-5568f3375d90 [Access date: 09.05.2020]

[14] Yeh M, Yan Z, Ulanova L, Begum N, Ding Y, Anh Dau H, Furtado Silva D, Mueen A and Keogh E 2016 Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View that Includes Motifs, Discords and Shapelets *IEEE 16th International Conference on Data Mining (ICDM)* (Barcelona) pp 1317-1322 https://doi.org/10.1109/ICDM.2016.0179

[15] Li Z, Wang K and He Y 2016 Industry 4.0-Potentials for Predictive Maintenance *6th International Workshop of Advanced Manufacturing and Automation* (Atlantis Press) pp 42–46

[16] Ansari F, Glawar R and Sihn W 2020 Prescriptive Maintenance of CPPS by Integrating Multimodal Data with Dynamic Bayesian Networks *Machine Learning for Cyber Physical Systems Technologies for Intelligent Automation* (Berlin: Springer Vieweg) pp 1–8

[17] Jasulewicz-Kaczmarek M and Gola A 2019 Maintenance 4.0 Technologies for Sustainable Manufacturing – an Overview *IFAC-PapersOnLine* **52**(10) 91–96

[18] Saxena A and Goebel K 2008 Turbofan Engine Degradation Simulation Data Set NASA Ames Prognostics Data Repository (http://ti.arc.nasa.gov/project/prognostic-data-repository) (NASA Ames Research Center, Moffett Field, CA)

[19] Zhu Y, Gharghabi S, Silva D F et al 2020 The Swiss army knife of time series data mining: ten useful things you can do with the matrix profile and ten lines of code *Data Min Knowl Disc* **34** 949–979 https://doi.org/10.1007/s10618-019-00668-6

[20] Shi J, Yu N, Keogh E, Chen H K and Yamashita K 2019 Discovering and Labeling Power System Events in Synchrophasor Data with Matrix Profile *2019 IEEE Sustainable Power and Energy Conference (iSPEC)* (Beijing, China) pp 1827-1832 https://doi.org/10.1109/iSPEC48194.2019.8975286

[21] Shakibay Senobari N, Funning G, Zimmerman Z, Zhu Y and Keogh E 2018 The Similarity Matrix Profile, an efficient method for detecting both low and high signal to noise ratio seismic events in very long time series *Poster Presentation at 2018 SCEC Annual Meeting*