Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 93 (2020) 1448–1453

www.elsevier.com/locate/procedia

53rd CIRP Conference on Manufacturing Systems

Intelligent Anomaly Detection of Machine Tools based on Mean Shift Clustering

Markus Netzer*, Jonas Michelberger, Jürgen Fleischer

Institute of Production Science (wbk), Karlsruhe Institute of Technology (KIT), Kaiserstrasse 12, Germany

* Corresponding author. Tel.: +49 1523 9502601; fax: +49 721 608 - 45005. *E-mail address:* markus.netzer@kit.edu

Abstract

For a fault detection of machine tools, fixed intervention thresholds are usually necessary. In order to provide an autonomous anomaly detection without the need for fixed limits, recurring patterns must be detected in the signal data. This paper presents an approach for online pattern recognition on NC Code based on mean shift clustering that will be matched with drive signals. The intelligent fault detection system learns individual intervention thresholds based on the prevailing machining patterns. Using a self-organizing map, data captured during the machine's operation are assigned to a normal or malfunction state.

Keywords: machine learning; manufacturing; machine tools; fault detection; anomaly detection

1. Introduction

Machine learning algorithms have led to significant breakthroughs in anomaly detection of machine tools [1-3].
Current state monitoring on machine tools depends largely on specific parameters with fixed values must be selected and examined by experts. The adaption to changing production processes and uncertain conditions requires expensive explicit re-specification of fixed intervention thresholds and monitoring through expert personnel. Moreover, in many real time machine tool processes, the required knowledge is not sufficiently available to deploy rule-based algorithms to meet the requirements of flexible and complex machine operations. Due to the increasing complexity of most modern machine tool use cases, autonomous systems for recognition and detection of anomalies are getting more and more important and, in many cases, mandatory [4].
Under these conditions a powerful anomaly detection is not satisfactorily applicable with respect to the current state of the art and justify the aspiration for autonomous self-learning fault detection systems. The main challenge for such an autonomous system is to learn which machine states are acceptable and which are not, especially when machine operations patterns are changing from time to time. Apart from a reliable accuracy, a further crucial requirement is an adequate run time efficiency, which facilitates a near real time fault detection under changing machine conditions.
Anomalies within the production process usually reflect quality defects in the underlying product or malfunction of the machine tool. To ensure the functionality of machine tools and to improve product quality, several data signals are available as data input for an anomaly detection system. The approach in this work focuses on the information stored in position, current and torque signals, but can be extended to further applications.
Based on a mean shift clustering algorithm, a new approach for an autonomous anomaly detection system is introduced. The combination of a density-based algorithm for the identification of recurring position patterns in combination with self-organizing maps for state capturing provides unique advantages

10.1016/j.procir.2020.03.043

and presents a novel approach for anomaly detection on machine tools.

2. Related work

The introduced approach to anomaly detection in this work is strongly interrelated with an autonomous cycle detection system which is based on a mean shift cluster algorithm. Apart from classic approaches with fixed fault detection limits predetermined by human, the research field on anomaly detection consists only of few works in the context of machine tool applications.

Fault detection approaches based on clustering are focused on the peer-to-peer comparison of machines as part of machine fleets [5] or not related to machine tool application [6-9].

Further works on fault detection that focus on varying operating conditions are based on unsupervised approaches and compare several distance measures [10,13].

First state-based and self-adapting approaches [11,12] are based on different kinds of state triggers which require little a priori knowledge about the concrete use case. Moreover, the handling of changing use cases such as varying production series is an unsolved challenge [11]. The approach presented in this work can handle with a higher degree of uncertainty and a lower information degree of the input data due to the advancement of a mean shift clustering algorithm.

Existing time series pattern recognition studies are mainly based on the early breakthroughs of Keogh [14,15,16] and Moen [15] and Lin [14,16]. These are the foundation of most following time series motif detection approaches and have a strong influence on the introduced combination of pattern recognition and anomaly detection.

The mean shift algorithm originates from the early works from Fukunaga et al. [17] who were the first to propose the basic clustering idea and introduced the term ‘mean shift’. The implemented and slightly modified algorithm is based on the non-blurring mean shift introduced by Comaniciu and Meer [18].

The Anomaly Detection Ensemble Approach (ADE) from Buda et al. (2017) [19] combines several state-of-the-art anomaly detection techniques not just to detect but also to predict anomalies on live data. Besides, ADE is coupled with a weighted window on incoming data to reward early detection. They employ different window weights to maximize various evaluation metrics, such as early detection, precision, and recall. They showed that adjusting these weights had the expected effect and that anomalies were on average detected 16 hours before they occurred.

The introduced intelligent fault detection system learns individual intervention thresholds based on the mean shift clustering algorithm. Using a self-organizing map, data captured during the machine’s operation is assigned to a normal or malfunction state. Apart from previous works this procedure is not relying on any knowledge about the machine operations on a machine tool and moreover, it can handle changing production series and patterns without human interaction.

3. Anomaly Detection

The main challenge for detecting anomalies for an autonomous machine learning based system is to learn which states are correct and which data fragments indicate a fault behaviour. This includes anomalies of the machine tool, the underlying operation process or the product, as long as they are represented in the available data signals.

To transfer this objective on a data problem the approach is divided into four major parts as shown in Figure 1.

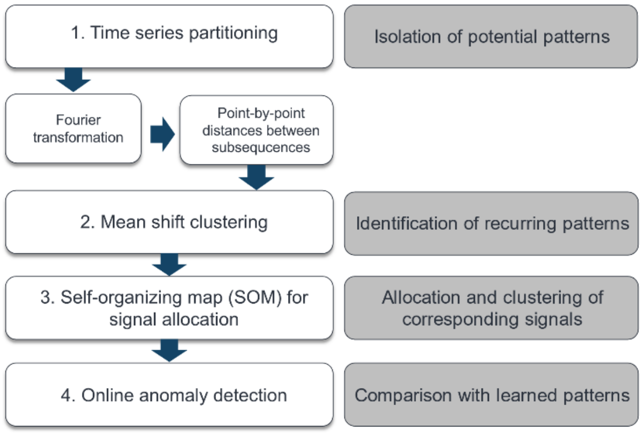

Figure 1: The system architecture relies on four subsequent data processing steps.

The time series segmentation aims to cut an arbitrary time series into logical distinct potential patterns. Opposed to Putz et al. (2017) [11] these are not event-based, but logical and consistent basic units which can be compared with each other. This comparison is performed through distance metrics and operates as input for the mean shift clustering algorithm.

Through the segmentation of sequences into clusters the algorithm seeks recurring patterns in the position data signal. The discovered recurring patterns are matched with the corresponding drive signals. To ensure that two recurring patterns with identical position signals but differing drive signals are recognized as different machine states, a self-organizing map (SOM) is used. The resulting machine states describe the recurring behavior of the machine tool and represent the learned information of the system and function as input for a subsequent anomaly detection.

Consequently, the system has learned relevant machine states and corresponding historic data signals as well as the individual intervention thresholds. This gained information provide a data base for the online comparison of incoming online live data streams of a machine tool with referenced signal patterns to detect anomalies. For the anomaly detection four methods for cycle detection (Simple Full Match Method, Delta Match Method, Delta Match Early Stop Method and Semantic Segmentation), five methods for cycle prediction (LSTM Networks, classification approaches like support vector machines, k-nearest neighbours, and decision trees), and three configurations for anomaly were presented. For each method, multiple different implementations with small variations were tested, resulting in a total number of over 240 system configurations excluding different signal clustering methods detection (Self Organizing Maps (SOM) and Agglomerative Clustering for drive signal clustering) and different datasets.

3.1 Time series partitioning

Time series partitioning denotes the partitioning of an individual time series into segments according to logical partition criteria. If the context of the use case is sufficiently available to the developer, an event-based trigger can be powerful as segmentation criteria [11]. But in many cases, such potential events are not a priori known. Then an arbitrary but consistent and topological partition criterion can be used to create time series subsequences as shown in Figure 2. Moreover, a combination of event-based and arbitrary topology criteria can be used to craft further information regarding to the machine tool data states.

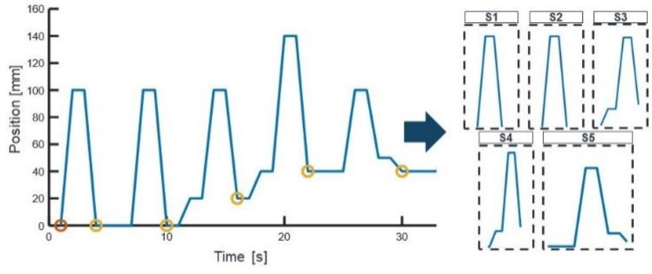

Figure 2: Time series partitioning into subsequences. In this case a local minimum search is used for partitioning.

Definition 1 (Time Series Subsequence):

A Time Series is a sequence $T = (t_1, t_2, t_3 \dots, t_n)$ *which is an ordered set of length n. A Time Series Subsequence is a subset of adjacent observations of a time series with length* $k \leq n$.

Contrary to pattern discovery approaches using a sliding window, the partitioning procedure in combination with clustering is not restricted by a fixed window length and focuses only on potentially relevant patterns rather than searching through all possible pattern compositions. As a result, it enables to reach lower run times without the use of data pre-processing approximation practises.

3.2 Mean shift clustering

To seek recurring time series subsequences and which are subsequently recurring patterns, a clustering approach is applied. For the comparison, a distance metric is needed which describes how similar the outlines of the position signal subsequences are. To enable full comparability a position offset adjustment of the subsequences is necessary. Therefore, a *Discrete Fourier Transformation (DFT)* of the subsequences is used to archive the following representation:

$$X_n = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi n}{N}k} \tag{1}$$

By removing the component $X_n = 0 = \sum_n x(n)$ for each subsequence, the offsets are removed, and the representation can be transformed back to a position-time signal. Afterwards, the subsequences are compared with each other by calculating pairwise the point by point distance of two subsequences each. To ensure comparability for patterns with differing lengths and to avoid distortion for patterns with a higher number of data points, a useful metric must fulfill the following definition:

Definition 2 (Comparable distance):

Independent from the amount of data points the comparable distance between two time series subsequences is described by the average point-by-point distance.

Following this requirement, a classic Euclidean distance is not useful because of the surrounding square root which applies a higher weight per point-by-point distance to shorter subsequences with less data points. Figure 3 illustrates the short coming of the Euclidean distance in a numerical example.

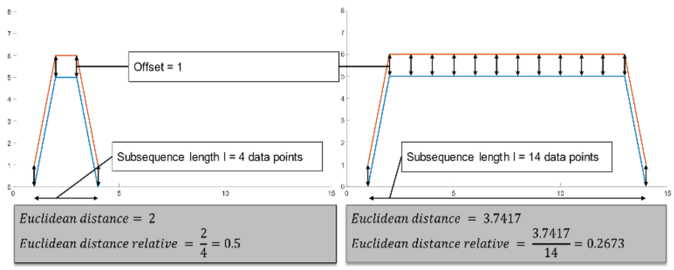

Figure 3: Numerical example showing that Euclidean distance cannot be used due to the interest in a distance metric relative to its underlying subsequence length *l*.

Depending on the expected dataset and the weighting of outliers, metrics that correspond to the following formula (2) can be applied to calculate a length-relative distance metric:

$$\frac{1}{l} \sum_{n=1}^{l} (| y_{i,n} - y_{j,n} | + 1)^p \tag{2}$$

Where
$l = length\ of$ compared pattern
$p = power$
$y_{i,n} = n - th\ data\ point\ y\ of\ i - th\ pattern$

The resulting pairwise distances between *n* time series subsequences are stored in a *n x n* distance matrix which functions as input for the mean shift cluster algorithm.
Because the number of recurring subsequence patterns is not known a priori a fix number of clusters can not be set as input parameter. Moreover, a clustering algorithm with predefined assumptions about the shape of the resulting clusters will lose information. Therefore, an algorithm is needed that combines these requirements together with the aim of a short run time and a low parameter choice complexity.

Mean Shift Algorithm

The *mean shift algorithm* is a hill-climbing algorithm that seeks modes of a density function using a generalized kernel approach. The algorithm is centroid-based and works iterative by updating centroid candidates to be the mean of the points within a given region [18]. Intentionally the mean shift algorithm was applied to cluster analysis in computer vision and image processing.

Algorithm 1 (Mean Shift)

Input: *Bandwidth parameter h*
Output: *Clusters with member points*
Filtering:
For $i = 1, \dots, N$ **do**
  Initialize $j = 1$ and $y_{i,j} = c_i = (x_i^s, c_i^r)$
  While not converged **do**
    Calculate $y_{i,j+1}$ according to

$$y_{i,j+1} = \frac{\sum_{i=1}^{n} c_i g\left(\left\|\frac{y_{i,j} - c_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{y_{i,j} - c_i}{h}\right\|^2\right)}$$

    $y_{i,j+1}$ *new position of the kernel window*
    number of points in the spatial kernel centred on $y_{i,j}$
  $y_{i,conv} = y_{i,j+1}$
  Assign $z_i = (x_i^s, y_{i,conv}^r)$
Segmentation:
For $i = 1, \dots, N$ **do**
  identify clusters $\{C_p\}_{p=1,\dots,P}$ of convergence points by linking together all z_i with distance $d = \frac{h_r}{2}$ from cluster centre where h_r is the radius of a cluster
For $i = 1, \dots, N$ **do**
  assign label $L_i = \{p | z_i \in C_p\}$
  eliminate spatial regions containing less than 2 subsequences

The bandwidth parameter represents the search radius and thereby the maximum distance between the cluster centres and the circumjacent data points [17]. Therefore, this parameter can be set as small as possible depending on the general white noise of the signal. With this the bandwidth can be directly linked to a *signal-to-noise ratio (SNR)* and implemented as adaptive autonomous parameterisation.

The computational complexity of the mean shift algorithm can be written as $O(Tn^2)$, where T is the number of iterations and n is the number of entries in the distance matrix. So, by doubling the amount of compared subsequences the run time increases by a factor of 4.

As opposed to Keogh et al. (2005) [14] this time series subsequence clustering approach is meaningful because the subsequences are not crafted using a sliding window with a fixed length and because of the missing specification for a fixed number of cluster centres which is a main advantage of the mean shift algorithm.

3.3 Self-organizing map for signal allocation

After receiving patterns from given time series, especially in position data, there must be an allocation of different time-corresponding signal parameters for example motor current or torque. Due to the matching of related signals the detected patterns can be found in online signals.

The mentioned allocation of drive signals to the detected pattern is represented by self-organizing map (SOM) clustering. Because of performance reasons, it is not appropriate to compare each online signal to all pattern-based drive signals in historical data. The SOM algorithm calculates the model m_i by using the given training data. Afterwards, it maps new data elements to a model m_i in the grid. The learning process has no need for labels during the learning phase (unsupervised) and therefore fits the requirements presented [20]. The unit space is a two-dimensional space whose length and width must be determined. First, the number of units n_u is guessed by $n_u = 5 * n_{samples}^{0,5}$ samples, where $n_{samples}$ is the number of samples provided. In a second step, the algorithm calculates the ratio between the length and the width of the unit space by using the two largest eigenvalues of the autocorrelation matrix of the input data to compute the ratio r as $r = \sqrt{e1/e2}$. To calculate the length and width of a hexagonal lattice from the ratio and the number of models, the algorithm applies the following formulas (3):

$$length = \sqrt{\frac{n_{models}}{ratio} * \sqrt{0.75}}$$
$$width = \frac{n_{models}}{length} \qquad (3)$$

After determining the size of the lattice, the model is trained. The result can be visualized as a so-called *umap* [20]. An example can be seen in Figure 4. Each hexagonal unit is represented on the lattice; the color depicts the *Euclidean distance* between the unit and its neighbors. The next step is to cluster the units in the grid by using a *k-means algorithm* for a range of k's starting from $k = \sqrt{n_{units}}$ running for five iterations. The best value for k is determined based on the sum of squared errors. Finally, the Davies-Bouldin Index and the Silhouette coefficient are calculated to evaluate the result.

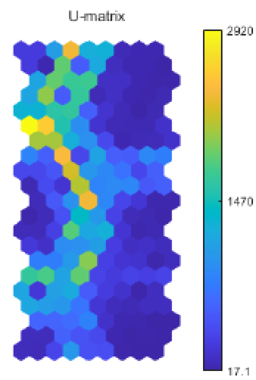

Figure 4: U-matrix consisting of hexagonal units.

3.4 Online Anomaly Detection

Anomaly detection simplifies to a comparative analysis between two signals, by using the clustered cycle signal data as well as position and probability of the current cycle. The process of anomaly detection can be seen in Figure 5. The process starts with a transformation of the representative and the online data; the next step is the calculation of the distance between representative and online data. This distance is then passed to the detection component, which decides whether an anomaly is present by using a threshold. Process input is extracted signal data from online data and for each signal, the corresponding representatives of the current model cycle. If the model cycle is ambiguous, representatives of each possible model cycle are used.

The anomaly detection component can handle an arbitrary number of signal transformations to base a decision on. One transformation implemented is the power spectrum; this transformation aims to aid the detection of oscillating anomalies that could not be discovered without it. Transformations are applied to each pair of online data and their corresponding representatives.

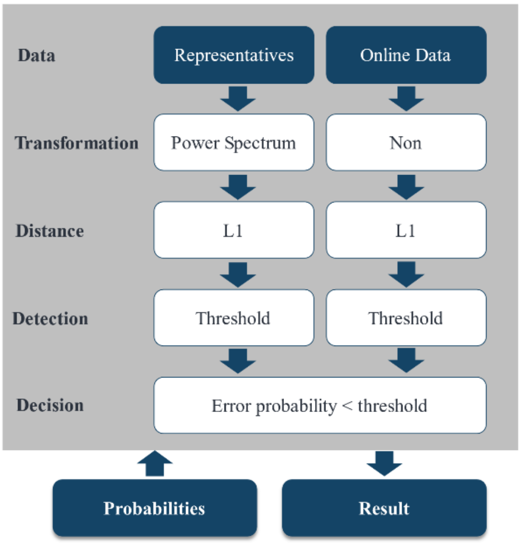

Figure 5: Online data anomaly detection procedure.

4. Experiments

The above mean shift-based cycle detection approach for anomaly detection is used in this study to verify its effectiveness, run time efficiency and transferability.

For this a drive signal data set is used which is generated by PLC signals input as shown in Figure 6. The milling pattern is described by several milling pockets in aluminium material using a conventional aluminium milling head. The aluminium block is prepared with irregularities produced by boreholes with differing diameters.

The data sampling rate of 500 Hz is chosen high enough to ensure correct representation of the signal patterns and to fulfil the Nyquist-Shannon-sampling theorem which is needed for meaningful Fourier transformation. Due to comprehension reasons the following observations are reduced to only one position axis and limited selected corresponding signals. The position signal of axis 1 serves as reference signal for the algorithm whereas the current and torque signals are assigned later after the mean shift clustering to represent the corresponding machine states.

To gain a maximum of information without a priori knowledge about the expected position patterns a combination of arbitrary and event-based partition criteria is used. The latter is a prominence-based peek search applied to the corresponding current signal of the milling tool. Thereby all contact points with the work material can be separated from idle position sequences. Within the relevant sequences the arbitrary partition criterion enables a further subdivision to enforce necessary precision of the later anomaly detection.

The used distance metric corresponds to the generic formula (2) with a power $p=2$ and thereby penalizing larger point-by-point distances between two position subsequences.

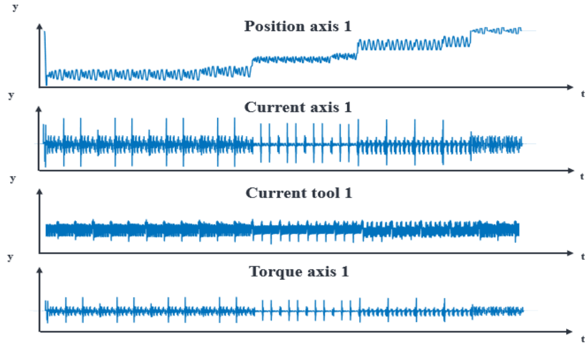

Figure 6: Underlying data set consisting of 520,228 data points for each signal. The position is used as reference signal, whereas current and torque are later allocated to identified subsequence patterns.

The bandwidth parameter is selected by an empirical non-linear model considering the standardized average signal-to-noise ratio and the peak prominence of the position data. As shown in Figure 7 the mean shift clustering result are recurring position patterns. In our experiments the mean shift provides reliable and highly accurate results for the identification of recurring position signal patterns.

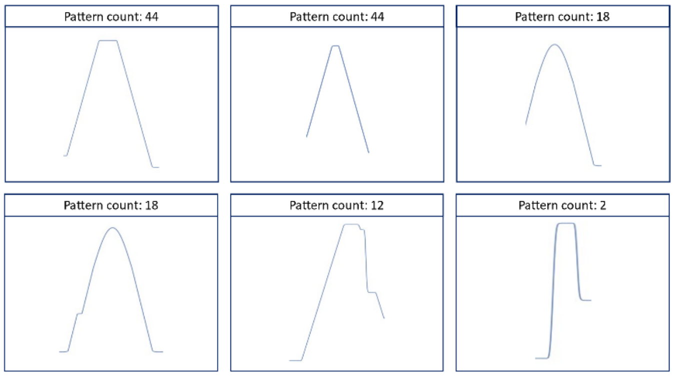

Figure 7: A small extract of the 95 identified clusters of recurring position signal patterns.

The experiments are conducted on an 8GB RAM laptop with quad core processor of 2.4 GHz. For the processing of the data set consisting of 520,228 data points and 712 subsequences the standard mean shift clustering implementation without pruning and boosting demands only 1.3 seconds. Therefore, run time evaluation shows that the approach is highly efficient and enables a near real time online learning of recurring patterns.

After the identification of recurring position signal patterns, the time-corresponding current and torque signals are allocated, and sub segmented using a SOM. In combination the signal patterns represent recurring machine operations and served as input for an online-anomaly detection. The latter analyses the incoming online position data stream to determine the current ongoing machine process and compare assigned current and torque signals.

As shown in Figure 8 the system detects the anomalies of the underlying current signal of tool 1 (milling head). The peak represents the irregularity within the aluminium material. The expected pattern of the current is learned and compared due to the recurrence of the position signal pattern.

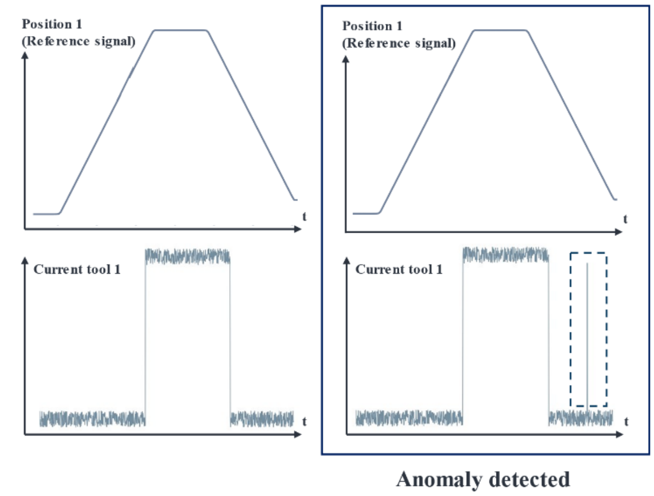

Figure 8: The system learns recurring position patterns and corresponding signals. An anomaly is detected due to the deviation of the assigned signal of the current of tool 1.

It must be mentioned that the displayed machine tool data show only a negligible noise behaviour for the position data. Further experiments disproved potential concerns that the clustering capability with noisy data is not appropriate. The validation is conducted on multiple datasets from different use cases of machine tools. Additionally, added white noise is used to constitute a worst-case scenario. Even in those cases the presented approach provided reliable results. The experiments show that the role of the bandwidth parameter selection gets more critical with an increasing noise level. Additional research on a refined use case specific bandwidth selection model could further increase the robustness of the presented algorithm.

Due to the softness of aluminium and the very slow processing speed of the experiments, an even higher peak can be expected in almost all machine tool operations. Further experiments on ordinary steel approved this proposition.

To reduce run time and complexity of the algorithm further investigation are necessary to effectively deploy the approach on more complex data signal combinations and therefor gain a higher information density.

Consequently, the presented approach experimentally verified its ability to work reliably under the most important dimensions of varying conditions. These include a priori unknown machine tool processes and tasks, varying noise level and varying product material.

5. Conclusions and outlook

The work introduces a new approach to enforce an autonomous anomaly detection system through a clustering-based pattern recognition algorithm. Contrary to other anomaly detection approaches, the system is capable to work without a priori knowledge about the expected machine operations and under uncertain, varying machine conditions. The algorithm learns a suitable parametrization and individual intervention thresholds based on the prevailing machining patterns. The experiments have proven that the mean shift-based anomaly detection system is highly powerful and efficient.

For further studies, we investigate on the application to broader time series problems including intelligent condition monitoring.

References

[1] Banerjee TP, Das S. Multi-sensor data fusion using support vector machine for motor fault detection. In: Information Sciences. Volume 217; 2012. p. 96-107.

[2] Purarjomandlangrudi A, Ghapanchi AH and Esmalifalak M. A data mining approach for fault diagnosis: An application of anomaly detection algorithm. In: Measurement. Volume 55; 2014. p. 343-352.

[3] Widodo A, Yang BS. Support vector machine in machine condition monitoring and fault diagnosis. In: Mechanical Systems and Signal Processing. Volume 21, Issue 6; 2007. p. 2560-2574.

[4] Lee J, Davari H, Singh J and Pandhare V. Industrial Artificial Intelligence for Industry 4.0-based Manufacturing Systems. In: Manufacturing Letters 18. 2018. p. 20-23.

[5] Lapira ER. Fault detection in a network of similar machines using clustering approach. Ph.D. Dissertation. University of Cincinnati, USA. 2012.

[6] Hou J, Xiao B. A Data-Driven Clustering Approach for Fault Diagnosis. In: IEEE Access PP. Volume 5; 2017. p. 26512 – 26520.

[7] Routray A, Rajaguru A and Singh S. Data reduction and clustering techniques for fault detection and diagnosis in automotives. In: IEEE International Conference on Automation Science and Engineering, CASE 2010. p. 326 - 331.

[8] Ferchichi SE, Zidl S, Laabidi K, Ksouri M and Maouche S. Meanshift Clustering Based Trend Analysis Distance for Fault Diagnosis. In: IEEE International Conference on Systems, Man, and Cybernetics, Manchester. 2013. p. 1877-1882.

[9] Liu S, Cao D, An P, Yang X and Zhang M. Automatic fault detection based on the unsupervised seismic attributes clustering. In: SEG 2018 Workshop: SEG Maximizing Asset Value Through Artificial Intelligence and Machine Learning, Beijing, China, 17-19 September 2018.

[10] Michau G, Fink O. Unsupervised Fault Detection in Varying Operating Conditions. In: Proceedings of the 2019 IEEE International Conference on Prognostics and Health Management, San Francisco. 2019.

[11] Putz M, Frieß U, Wabner M, Friedrich A, Zander A and Schlegel H. State-based and Self-adapting Algorithm for Condition Monitoring. In: Procedia CIRP. Volume 62; 2017. p. 311-316.

[12] Frieß U, Kolouch M amd Putz M. Deduction of time-dependent machine tool characteristics by fuzzy-clustering. In: Beyerer J., Kühnert C., Niggemann O. (eds) Machine Learning for Cyber Physical Systems. Technologies for Intelligent Automation. Volume 9; 2018. p. 7-17.

[13] Zhang R, Tao H, Wu L and Guan Y. Transfer Learning with Neural Networks for Bearing Fault Diagnosis in Changing Working Conditions. In: IEEE Access PP. 1-1. Volume 5; 2017. p. 14347-14357.

[14] Keogh E, Lin J. Clustering of time-series subsequences is meaningless: implications for previous and future research. In: Knowledge Information Systems 8. 2005. p. 154–177.

[15] Keogh E, Mueen A. Finding Repeated Structure in Time Series: Algorithms and Applications, Vancouver. 2015.

[16] Lin J, Keogh E, Lonardi S, Patel P. Finding Motifs in Time Series. In: Proceedings of the Second Workshop on Temporal Data Mining. 2002. p. 53-68.

[17] Fukunaga K, Hostetler LD. The estimation of the gradient of a density function, with application in pattern recognition. In: IEEE Trans. Information Theory. Volume 21, Issue 1; 1975. p. 32–40.

[18] Comaniciu D and Meer P. Mean shift: A robust approach toward feature space analysis. In: IEEE Trans. Pattern Analysis and Machine Intelligence. Volume 24, Issue 5; 2002. p. 603–619.

[19] Buda TS, Assem H and Xu L. ADE: An ensemble approach for early Anomaly Detection. In: 2017 IFIP/IEEE Symposium on Integrated Network and Service Management. 2017. p. 442–44

[20] Kohonen, T. (2001), Self-organizing maps, 3rd ed.. Aufl., Springer series in information sciences, 0720-678X, Bd. 30, Springer, Berlin and London