



Study on learning efficient stroke representations in clocked sheet metal processing: theoretical and practical evaluation

Philipp Niemietz¹ · Marek Fencł¹ · Thomas Bergs^{1,2}

Received: 5 October 2022 / Accepted: 7 January 2023 / Published online: 2 February 2023
© The Author(s) 2023

Abstract

Clocked manufacturing processes such as sheet metal forming and cutting processes pose a challenge for process monitoring approaches due to inaccessibility of tool components and high production rates which make direct measurement of the physical process conditions unfeasible. Auxiliary data such as force signals are acquired and assessed, often still relying on control and run charts or even visual control in order to monitor the process. The data of these signals are high-dimensional and contain a large amount of redundant information. Therefore, the processing of such signals focuses on compressing information into as few variables as possible that still represent the important information for the manufacturing process. Due to repeatability in clocked sheet metal processing, the data generated consist of a series of time series of the same operation with varying physical conditions due to wear and variations in lubrication or material properties. In this paper two major research objectives are identified: (i) *the theoretical evaluation of representation learning methods* in context of clocked sheet metal processing, and the connection with (ii) *the practical evaluation of the learned representations* with a given use case to track the wear progression in series of strokes. The contribution of this paper is the comparison of varying time series representation learning techniques and their performance evaluation in a theoretical and practical scenario.

Keywords Fine-blanking · Force signal · Time series data mining · Representation learning · Wear

1 Introduction

Clocked sheet metal processing technologies are complex, efficient and used for the mass production of in part safety critical set of workpieces in automotive and aerospace industry [1]. The constant pressure to reduce cost and improve quality in these industries enforces an improvement of the process capabilities. At the same time, more digitization is required in order to enable networked production. However, extracting relevant data in clocked sheet metal processing is challenging and the inaccessibility of tool components

makes direct measurement of physical process conditions hardly feasible. Nevertheless, the changes in physical process conditions have major effect on the process stability and need to be determined indirectly. It has been shown that acoustic emissions [2], acceleration [3] and force signals contain valuable information about wear progression, fault diagnosis, and process stability in general. The force signal of a single stroke, e.g. a forming or blanking operation, is a periodical time series that remains almost identical in a series of strokes with the same tool setup. Minor deviations or changes of the profile can represent changes in the underlying physical conditions of the process [4].

The signal is a high-dimensional time series in which every observations or data point is viewed as one dimension. Generally, when dealing with such high-dimensional raw data, two main problems occur — high computational costs for processing and analysing them and a need for large storage space. Therefore, a representation or an abstraction of the data is required to process the data effectively or even online. While reducing the dimensionality of a time series, the aim is to preserve as much information contained in the

✉ Philipp Niemietz
p.niemietz@wzl.rwth-aachen.de

Marek Fencł
m.fencł@wzl.rwth-aachen.de

Thomas Bergs
t.bergs@wzl.rwth-aachen.de

¹ Laboratory for Machine Tools and Production Engineering (WZL) at the RWTH Aachen, 52074 Aachen, Germany

² Fraunhofer Institute for Production Engineering IPT, 52074 Aachen, Germany

raw data and as much fundamental characteristics of the time series as possible.

In an early work of Jin et al. [5], a feature preserving data compression method was developed on strain (tonnage) signals of a stamping press. The diagnosis of stamping processes showed that Principal Component Analysis (PCA) applied on low-sampled force data is already capable of extracting some useful information for detecting coarse variations. In [6], PCA was used as a statistical feature extraction method and its capability to explain process variations was compared with features engineered by an expert in field. It has been shown that features extracted by PCA perform very similar on the task of identifying variation in the process. Engineered features were more interpretable, but PCA application required far less domain knowledge and is also transferable to different kinds of cutting processes [6].

Niemietz et al. [7] demonstrated a potential of detecting a wear increase using an autoencoder (AE). A convolutional AE (CAE) was trained on the first strokes of the experiment and its reconstruction error was investigated as a measure for tool wear increase. In a recent paper, Niemietz et al. [8] investigated the force signal variations in clocked sheet metal processing by applying a specially designed feature extraction toolkit to different sub-segments of the process signature and projected the features onto the actual wear progression, utilizing different techniques such as PCA or an AE. For selected non-corrupted data sets the extracted variation was successfully linked to the wear progression.

Wear progression is one major but not the only factor influencing the signals in a clocked-sheet metal processing. A manufacturing process is regarded as stable only when all the measured variables have constant means and constant variances over time [9]. This requirement seems obvious in order for the process to perform reliably and predictably over time. Widely used and well-established tools for stability monitoring are run charts and control charts. To enable digitization and drive innovation forward, new approaches to stability assessment are researched. Zhang et al. [10] described a statistical process control monitoring system that integrates outlier identification, using clustering and a feature extraction technique Discrete Wavelet Transform (DWT). Recently, Jensen et al. [11] introduced the stability

index (SI) which they argued was the best practice for stability assessment.

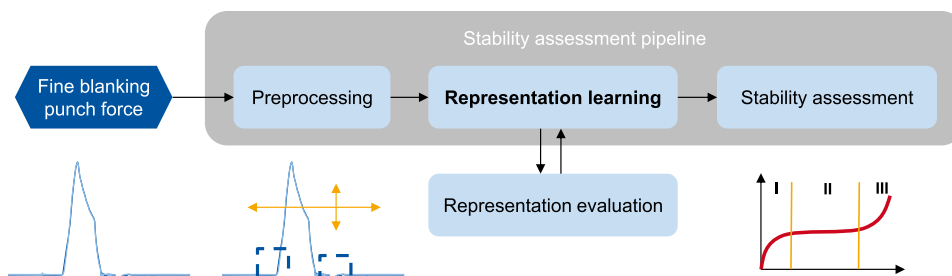
In general, the majority of the research has been conducted by applying one representation learning or dimensionality reduction method on a specific data set. In this paper, the contribution is a unique comparison of a range of widely used representation learning methods by an evaluation with different sets of field data acquired in clocked sheet metal processing technologies. The aim is specifically to a) *identify and evaluate time series representation learning methods in context of an exemplary clocked sheet metal processing technology*, since there was generally no best time series representation learning method presented in the literature up to this date, and to b) *explore new possibilities for unsupervised assessment of the stability of the representations* for building a novel signal stability assessment pipeline (see Fig. 1). The stability of the representation is then linked to the progression of wear during a series of strokes, and therefore, a step to link the stability of a signal to the stability of the process is given. Fine-blanking is investigated as an representative of the clocked sheet metal processing technology.

This paper is organized as follows: In Chapter 2, methods and approaches utilized throughout the paper are described. The following Chapter 3 first introduces the data analyzed, the way they were acquired and the different representation learning methods are evaluated afterwards. The theoretical approach is linked with a practice of tool wear monitoring and process stability in Chapter 4. The last Chapter 5 then concludes on the results and findings and proposes the possible next steps.

2 Methodology

Time Series Representation Learning. The aim was to implement a range of time series representation learning methods that covers all the known approaches and levels of complexity and analyze the outputs. This part provides an overview of different representation learning methods ranging from already established Discrete Fourier Transform (DFT) to novel ones such as AE.

Fig. 1 Scheme of the different areas of the stability assessment pipeline development with an emphasis on the efficient representation learning part



DFT decomposes an input signal into a series of sinusoids which characteristics can be easily measured. Proposed by Agrawal et al. [12] for sequence databases mining, DFT nowadays belongs to the major time series representation techniques [13]. A time series is mapped to lower dimensionality space by using only the most significant Fourier coefficients. As opposed to DFT, DWT considers localization of the identified frequencies in the time domain. It uses wavelets (functions) constructed from a “mother” wavelet by translations and dilations.

As opposed to other presented representation techniques, Piecewise Aggregate Approximation (PAA) is a method designed specifically for time series by [14]. PAA divides the time series into ω equisized windows and calculates an average in each of the windows. The established method used in the literature to extract the relevant variations from the signals is the PCA. PCA decomposes a data set to orthogonal components that explain maximum amount of variance. These, so called principal components, are linearly uncorrelated variables of which the first few are usually able to explain most of the variance. The idea of normal PCA method is to remove second-order dependencies from the data. PCA could thus produce misleading results when higher-order dependencies exist between the variables. To solve this, data can be transformed to a more appropriate higher-dimensional space using a nonlinear function (i. e. kernel) and afterwards projected back to lower-dimensional subspace. This approach is termed as Kernel PCA (KPCA).

Recently proposed and promising representation learning techniques are often based on machine learning. This holds true for a technique utilizing an AE for condition monitoring purposes, introduced by [15]. First part (encoder) can be viewed as a classical fully connected neural network which compresses the input signal x into a low-dimensional representation (embedding). However, as opposed to a simple neural network, an AE has a second part (decoder) decoding the representation with a goal to reconstruct the input signal as accurately as possible to the output signal \hat{x} . The encoding (low-dimensional representation) from the narrowest part of the whole architecture (i. e. bottleneck) should preserve all the information needed for the AE to reconstruct the signal again.

Another time series representation method, recently introduced in [16], addresses the problem of insufficient research efforts in representation learning of time series. Generic RepresentAtlon Learning (GRAIL) is similarly to KPCA, based on nonlinear dimensionality reduction principle (kernels). The kernel function parameters are however estimated in more sophisticated way. In [16], the implemented kernel is called Shift-Invariant Kernel (SINK).

The last representation technique investigated in this paper is Time Series Feature Extraction Library (TSFEL). Introduced by [17], the aim of this Python package is to transform a time series into a set of properties (features) which characterize the time series (feature space).

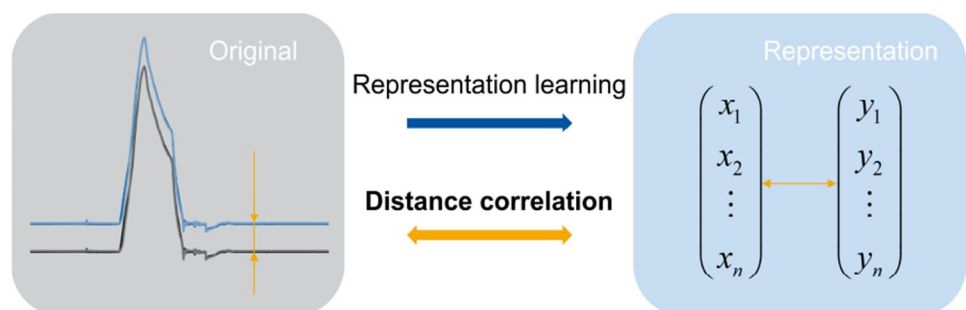
Representation Learning Evaluation Techniques. A desired representation learning method should:

- Reduce dimensionality,
- Shorten computational time,
- Preserve local and global characteristics of the data.

To compare representation learning methods based on these criteria, four different evaluation techniques were designed or implemented. As opposed to most of representation evaluation done in the research (e. g. in [16]), there are no labels available in this use case which posed a challenge in the design of the techniques introduced in the following.

The first technique focuses on the global characteristics of the data and uses correlation of time series distances. Due to the preprocessing steps and exact alignment of process signatures, Euclidean distance is applied to measure distances between the time series. At first Euclidean distances between consecutive original time series are measured. Next, representations of the original time series are computed using the representation method which performance is to be evaluated (see Fig. 2). Once all data are transformed to the low-dimensional representation space, Euclidean distance measure is applied again to the consecutive representations. The Pearson correlation coefficient is than used to measure the strength of linear correlation between the Euclidean distances in the original space and the representation space. The distances between two consecutive time series in the original space and the representation space should preferably

Fig. 2 Distance correlation evaluation technique schema



stay the same and the correlation should be high for the desired representation.

Time series classification has been used to evaluate representations in the majority of papers concerned about time series representation learning. The basic idea is to assign each time series sample to a class which is in the case of 1-nearest neighbor (1-NN) classification the class of the one nearest neighbor. Time series, respectively their representations, are classified using leave-one-out 1-NN. At first, the original time series are labeled with their indexes and classified (each time series is assigned the label of its first nearest neighbor). Next, representations are obtained, and the same procedure is repeated with them. Finally, the accuracy is calculated as a score of classifying the time series in original space with the same label (index) as the representation in the representation space. In other words, the nearest neighbor of a time series in the original space is ideally the same one as the nearest neighbor of its representation in the representation space. Since the process signatures are very similar, expecting exactly the same label would be very strict. For this reason it is considered a match when the labels (indexes) fall within a toleration window of ± 10 . High accuracy is in this case an indicator of an ability of the representation learning method to preserve local characteristics of the data.

Different so-called measures of distortion were examined in [18] in order to assess the quality of a data representation in context of machine learning applications. Let X and Y be arbitrary finite metric spaces. Given original distance $d_X(Q, C)$ of two time series Q and C , and a distance of representations $d_Y(f(Q), f(C))$ of the same time series, a distortion measure is a summary statistic of the pairwise ratios $\rho_f(Q, C)$ (see Eq. 1). Intuitively, the distortion is small when the ratios of distances $\rho_f(Q, C)$ is close to 1.

$$\rho_f(Q, C) = \frac{d_Y(f(Q), f(C))}{d_X(Q, C)}. \quad (1)$$

Based on established sigma-distortion measures, Vankadara et al. [18] designed their own distortion measure which evaluated specific requirements for later machine learning application.

Finally, to ensure maximal possible dimensionality reduction, intrinsic dimension is determined as a minimal number of features needed to represent a data set with minimal information loss. The intrinsic dimension estimators have gained considerable attention in the last years because of their relevance for data reduction. Ceruti et al. [19] introduced their Dimensionality from Angle and Norm Concentration (DANCo) algorithm for intrinsic dimension estimation which was used in this paper. For the evaluation purposes, intrinsic dimension ratio is defined as the intrinsic dimension of a representation divided by the number of its dimensions.

Stability Assessment Methods. Out of a variety of process monitoring approaches applied throughout literature, three best suited methods were implemented and modified for the specific task of clocked sheet metal processing technologies force signal stability assessment. The first and most obvious characteristic of a stable process is a mean that remains constant over time. The Analysis of Variance (ANOVA) method proposed by [20] aims to detect significant changes of the process mean. In order to use the ANOVA method, data are divided into subgroups and a null hypothesis that all subgroup means are equal is stated. Counter intuitively, it compares subgroup variation to within subgroup variation to assess the mean differences. To test the null hypothesis, F test needs to be conducted with the F value computed defined in Eq. 2 [9].

$$F \text{ ratio} = \frac{\text{variation between subgroup means}}{\text{variation within the subgroups}}. \quad (2)$$

In Fig. 3, the design of a modified ANOVA approach (moving ANOVA) for stability assessment is illustrated. At first, individual dimensions of all representations are merged into corresponding series. In this way, e. g. first dimensions of all representations form one series and can be used as an input for ANOVA. With 10-dimensional representations, 10 input series exist. In the next step, the F ratio is calculated for every set of 20 consequent data points sampled from the series. With an offset of the first and the last 10 data points, every point has now its F-value characterizing behavior of the series in its closest neighborhood. To get the final stability assessment output, F value series of all the dimensions are simply averaged into the resulting plot. This output is an indicator of stable and unstable parts in the signal and can be used for the stability assessment.

The second method relies on assessment of differences in variations. The idea behind SI from Jensen et al. [11] is that the long-term variation captures random as well as non-random patterns in the data (i. e. drift). The goal of the short-term variation is as opposed to the long-term one to only capture the random patterns (e. g. common causes of variation). If the process is stable, the long- and short-term variation are similar and SI is close to 1. If there are non-random patterns in the data, the ratio will be greater 1. For purposes of this paper, local stability assessment is of interest and sets of 50 data points are therefore sampled from the data. A standard deviation of a set is utilized as the long-term variation estimate. The 50 data points are then divided into 10 subsets à 5 data points which are used for the short-term variation estimate.

Finally, the regression analysis identifies trends throughout the signal. All of upward or downward shifts, increasing or decreasing trends could be interpreted as signs of instability. Possible drift or trend can be detected by fitting a straight

Fig. 3 Moving ANOVA approach—extracting dimensions from representations, ANOVA of every set and finally average over sets throughout all dimensions

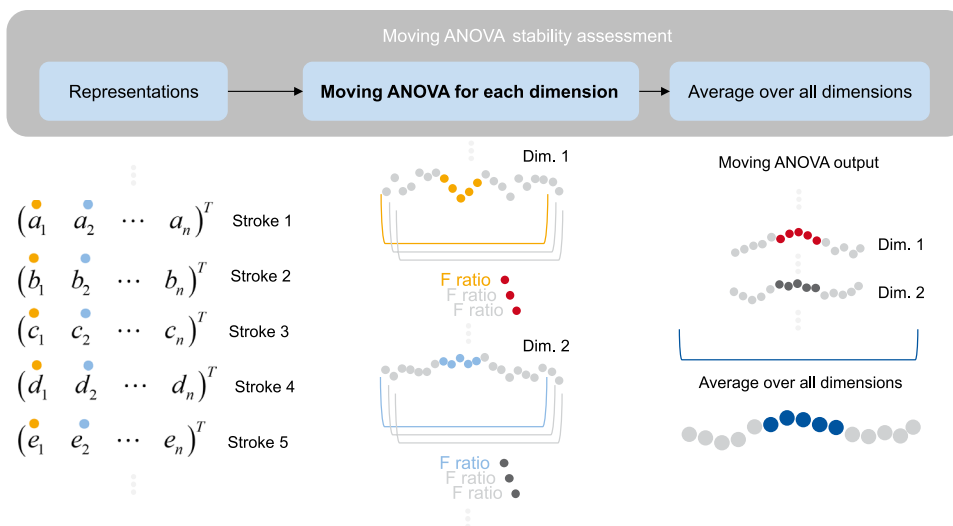


Table 1 Specifications of the used data sets

Data Set ID	Strokes	Material	Wear data
E_2	3204	X5CrNi18-10	Yes
E_4	3340	X5CrNi18-10	Yes
KS_5	8112	51CrV4	No
WZV_c	29,142	100Cr6	No

line to the data — linear regression. The null hypothesis is in this case that all of the regression coefficients are equal to zero. This would imply that the model used has no predictive capability. By running a F-test, it is inspected if the fitted line has a statistically significant slope or not. If the null hypothesis is rejected and the slope is significant, it can be concluded that the signal is unstable [9].

3 Representation learning methods assessment

Data Sets.

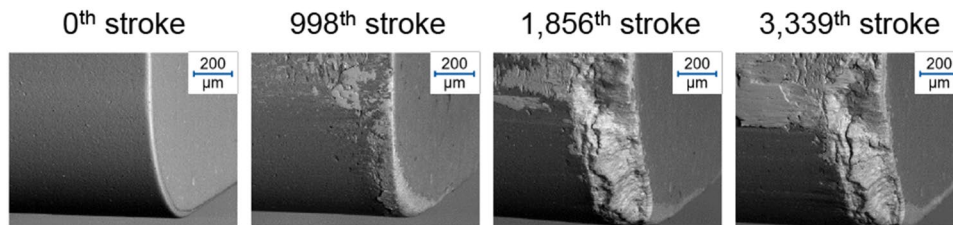
For this study, a selection of four experiments is presented in Table 1. The general setup of each experiment similar, utilizing a servo-mechanical Feintool XFT 2500 Speed

fine-blanking press with a straightener to relieve the residual stresses from the metal sheet [1], and a lubrication system to apply a lubricant film. All experiments were conducted with a rate of 50 strokes per minute and a metal sheet thickness of 6 mm. The original experimental data came from a series of trials with industry collaborations. A number of relevant data sets were selected for this study to illustrate the applicability of the approach presented.

Four piezoelectric sensors were placed at the punch, acquiring data with a rate of 10 kHz. Raw force signal data from the sensors are long time series of sensor profiles (strokes) repeating over time. These signals were preprocessed-segmented into individual strokes and cleaned-in the same way as in [8]. These preprocessing steps are crucial for further processing and analysis.

Since tool wear is one of the major factors influencing the stability of the process [7], microscope images of the tool cutting edge have been taken in predefined intervals for experiment E_2 and E_4 in order to observe and assess the wear progression during the experiment (see Fig. 4). For each examination of the wear progression, the press was stopped, and the fine-blanking tool was disassembled. The data representing the wear increase has been generated by approximating the amount of damaged coating visible through the SEM images. To do so, image processing techniques were used to isolate the area of damaged coating and based on the

Fig. 4 SEM images of tool wear taken in regular intervals from the start until the end of E_4 experiment



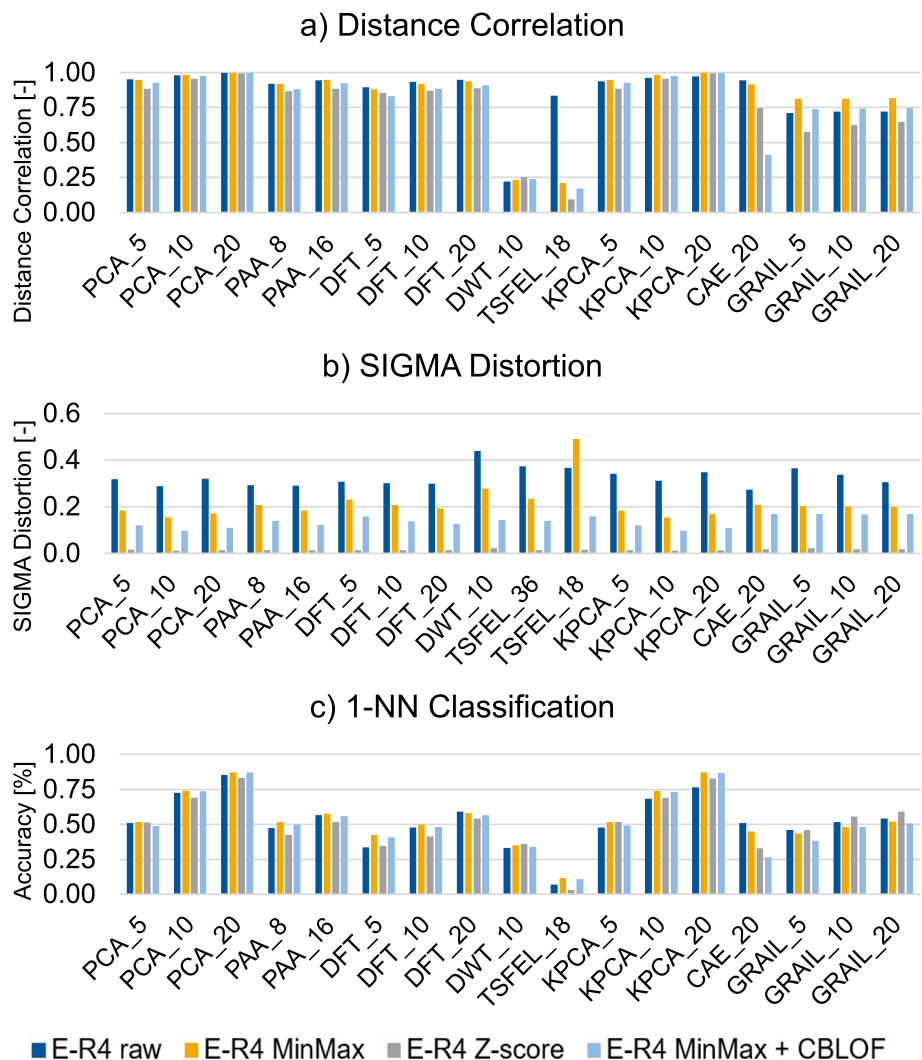
number of pixels of the total image, the share of damaged tool surface of the total image is taken as a measure for the wear in this study. For details see [8].

Preprocessing Influence on Representation Learning and Evaluation Methods. In order to assess the representation learning methods comprehensively and to decide which of the introduced methods performs the best on the fine-blanking force signals, effects of different preprocessing steps on the methods need to be investigated. Four different preprocessing variants are chosen to investigate their influence on the representation learning and evaluation methods. In case of raw data, the only preprocessing step taken on the cleaned and segmented data is the combination of the 4 punch force sensor signals into a single one. This is done simply by averaging the signals and leaving the potentially corrupted signals out. For normalization and standardization steps, MinMax scaling—normalizing each data point of each stroke to lie between 0 and 1—and the so called Z-score—standardizing the data such that their distribution

will have a mean value of 0 and a standard deviation of 1—were used. In order to investigate the effects of outliers (anomalies) in the data on the representation and evaluation methods, the Cluster-based Local Outlier Factor (CBLOF) anomaly detection is used to identify outliers which are afterwards removed. The dimensionality of representations is varied as well with the aim of finding an optimal dimensionality for each method and corresponds to the number following the name of the method in Fig. 5.

Results are shown in Fig. 5. The distance correlation results for different representation learning methods are not affected by the preprocessing in greater extent, except of DWT, TSFEL, CAE, and GRAIL. These four methods also perform on average worse than the rest evaluated by distance correlation. Normalization and combination of normalization and anomaly detection improve the distance correlation slightly for all the well performing methods. Dimensionality does not have greater impact, with results becoming rather better with increasing representation dimensionality.

Fig. 5 Performance of the representation learning methods of different dimensionalities (the number following the name of the method) measured by the different evaluation learning methods depending on the preprocessing steps evaluated by **a** Distance Correlation, **b** Sigma Distortion and **c** 1-NN Classification



Sigma-distortion pointed out the importance of the preprocessing steps selection. While DWT and TSFEL representations show slightly worse results than other methods overall, the prominent differences originate from the preceding preprocessing. The sigma-distortion values are minimized by Z-score standardization preprocessing since it takes the distribution of the data into account. MinMax scaling as well as anomaly detection improve the resulting distortion.

Finally, 1-NN classification gives a clearer picture about the representation learning methods. The sophisticated GRAIL or untuned CAE produce accuracy comparable with the simplest PAA. PCA and KPCA perform already at 5-dimensions superior to other methods, followed by DFT, PAA, and GRAIL. Preprocessing does not significantly affect the results in case of 1-NN classification.

Additionally to the evaluation methods introduced in previous part, elapsed time is measured during the run of each representation learning method in order to assess its computational cost. The averaged time needed for the different representation learning methods to run varied significantly with CAE being clearly the most expensive, followed by GRAIL and TSFEL.

Interim Conclusion. It was shown that the normalization as well as the anomaly detection preprocessing step improve performance assessed by the evaluation methods in most cases. The CAE implementation is computationally expensive compared to the rest of the methods and would require further tuning efforts in order to unfold the potential of this powerful approach. TSFEL and DWT implementations show bad results in most of the evaluation cases. These three methods are therefore not further examined.

Influence of Different Data Sets on the Representation Learning Methods Performance. The ultimate objective of this paper is to design a robust and efficient stroke representation learning method independent of the data set at hand or even the analyzed process. It is therefore desirable that the representation learning methods perform as consistent as possible throughout the whole range of data sets. Three representatives of different lengths and varying specifications are used. All of them were cleaned, segmented, and normalized using MinMax scaler with anomalies removed at the end.

In Fig. 6, the results are shown. The distance correlation results have similar trends across the data sets except of the GRAIL method results, which depend on the length of the data set significantly.

Sigma-distortion is in absolute numbers lower for the KS_5 and WZV_c data sets. The reason for this effect could be that the longer data sets have distributions closer to normal distribution. Application of the Z-score as a preprocessing step might eliminate this effect. Otherwise, the trends are similar between the representation learning methods with

the exception of KS_5 , where the PAA and DFT dimensionality have the opposite effect on the sigma-distortion.

1-NN classification as well as intrinsic dimension ratio show consistent results across data sets and confirm the superior performance of PCA. Intrinsic dimension ratio gives a good overview of the optimal representation dimensionality. The trends in the results are similar for all the representation learning methods and data sets. Based on this evaluation method, the optimal number of dimensions lies somewhere between 5 and 10 and is nearly independent of the preprocessing.

As for the time complexity, KPCA and GRAIL scale quadratically with the amount of data, while for example 5-dimensional GRAIL taking 30-times more time to run than 5-dimensional KPCA. Other methods are considerably less expensive and scale linearly.

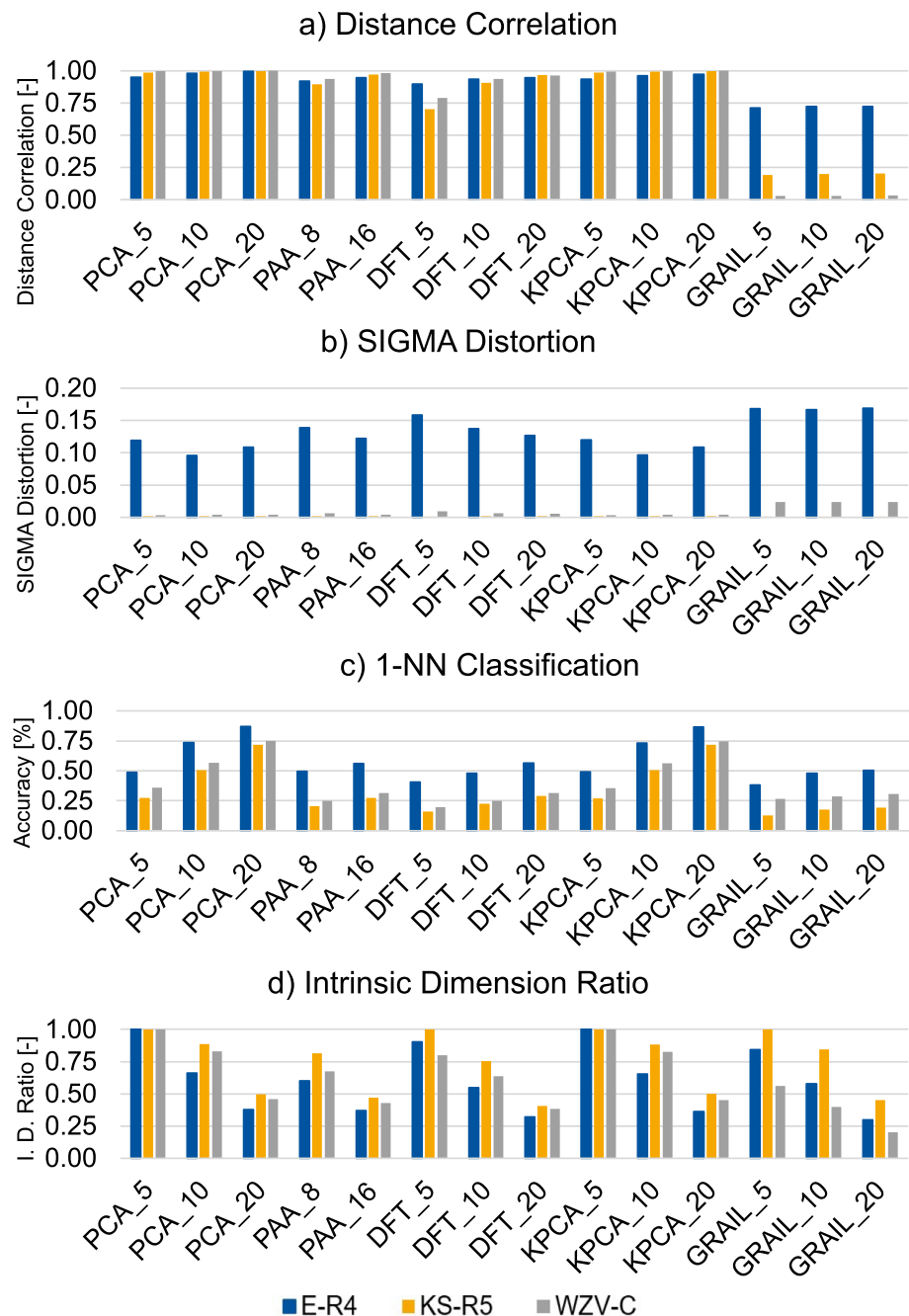
Discussion. Some of the evaluation methods show different trends for PAA and DFT dimensionality between the different data sets and a closer investigation is required in order to understand the causes of this behavior. PCA and KPCA show the best results in the distortion measures and 1-NN classification across data sets, followed by DFT and PAA. KPCA results are nearly identical with those of PCA and thus do not bring any new extra valuable insights which would justify the higher computational cost. GRAIL results do not justify their computation expenses. In fact, GRAIL performs worse than other representation learning methods in most of the evaluation points of view.

PCA and DFT perform the best on the use case fine-blanking based on the quality of their representations evaluated in an unsupervised manner. These methods show promising results and are primarily applied in the next part in order to generate stroke representations.

4 Exploration of signal stability assessment and monitoring of condition changes

The main objectives of this part are the practical application of the considered approaches and an exploration of a quantitative evaluation of the presented representations. It is to investigate if the identified signal stability assessment methods can capture the trends towards stable/unstable process state based on the representations and which of the methods can also do it consistently for different data sets. For all evaluation data sets all process parameters have been kept constant during the execution of the experiment. Therefore, the tool wear can be considered to have the largest impact on the process stability, and according to the hypothesis of this paper, to the stability of the representations of strokes over time. According to Behrens et al. [21], three stages of wear rate occur during the lifetime of a tool. An unstable run-in period at the beginning, followed by a more stable low wear

Fig. 6 Performance of the representation learning methods on different data sets evaluated by **a** Distance Correlation, **b** Sigma Distortion, **c** 1-NN Classification and **d** Intrinsic Dimension Ratio



rate stage, and the last stage of accelerated wear until a possible breakage of the tool and the end of the tool's life-cycle. Since all the data sets were acquired during experiments starting with an unused punch and executed over several thousand strokes, the punches are expected to have reached the beginning of the second wear phase roughly by the end of each experiment.

This chapter is organized as follows: the three stability assessment methods, moving ANOVA, SI and regression analysis, are applied on 5-dimensional PCA to qualitatively evaluate the ability of representations to capture tool

wear progression based on three experiments. The combination of normalization and standardization is applied as preprocessing while the removal of anomalies leads rather to an unnecessary data loss than an improvement and is thus not further considered. Finally, the tool wear was measured during the acquisition of E_2 and E_4 data sets which allows for a quantitative evaluation of the representation learning methods.

E_2 and E_4 Data Sets. The data were acquired continuously only between tool inspections (in case of E_2 they took place around 800th and 1,800th stroke).

The normalized evolution of the maximum punch forces during the E_2 experiment is shown in Fig. 7a) (along with the results of the stability assessment methods). Tool inspections were correctly identified as the cause of process instability by all stability assessment methods. These events occurred intentionally and are followed by an unstable ramp-up period of the entire machine/process. However, other periods of instability identified are not related to tool inspection and could have a variety of possible causes. Tool wear is assumed to be the main cause, with an expected evolution of the wear rate corresponding to the gradient of the normalized cumulative stability measures, which is higher at the beginning and decreases during the experiment (except for the unstable events).

Moving ANOVA is more sensitive to the sudden events and the SI performs similarly with the regression analysis for both, the E_2 and E_4 data sets.

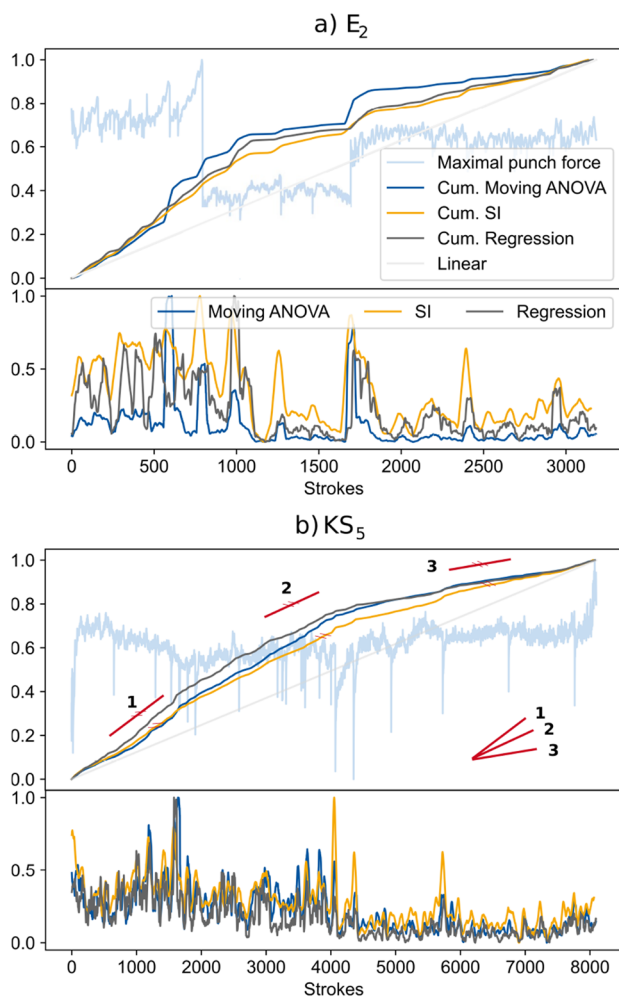


Fig. 7 Normalized maximal punch forces of raw data and stability assessment pipeline output of **a** E_2 and **b** KS_5

KS_5 Data Set. In the case of KS_5 data set, all three stability assessment methods indicate that the signal was more unstable in the beginning (gradient in the period 1 in Fig. 7 b) and converged towards more stable stage (period 2 and 3) until around the 7,500th stroke where it began to become unstable again. The scanning electron microscope (SEM) images in Fig. 4 also confirm the results of the stability pipeline outputs and show, that the wear rate was considerably higher in the first 5,000 strokes and decreased in the second half of the experiment. Minor differences between the stability assessment methods can be observed in the case of KS_5 data set. The regression analysis identifies slightly higher instability at the beginning, resulting in its higher gradient in cumulated form. The SI is the most sensitive to sudden instability events.

Wear RMSE Evaluation. Tool wear measurements during acquisition of E_2 and E_4 data sets enable a final evaluation of the representation learning methods in context of the stability assessment techniques and wear progression. In the lower part of Fig. 8, a polynomial line was fitted to the wear values. Curves of cumulated moving ANOVA, SI as well as regression stability are shown and their abilities to approximate the wear progression are examined. To quantify the fit to the wear progression, the Root Mean Square Error (RMSE) of each line is calculated (Fig. 8 upper part).

The resulting plot proves that wear can be approximated well by analysis of the differences in signal variations (SI) for E_4 data set. In case of E_2 , the stability assessment technique ability to capture wear progression depends significantly on the representation learning method. Moving ANOVA (differences in means) potentially identifies different causes of instability in the data. DFT representations seem to preserve information about the wear progression consistently throughout data sets, whereas TSFEL as a pure feature extraction method performs poorly. GRAIL approximated the wear progression of E_4 data set well but this was not consistent with its E_2 results.

Discussion. A possible link between theoretical evaluation and practical use for condition monitoring purposes with quantitative evaluation was shown. As for the signal stability assessment, the capability of the stability pipeline highly depends in the data it is applied on. For E_2 and E_4 data sets, the SI and the regression analysis on PCA as well as on DFT resulted in interpretable and convincing output corresponding with the tool wear measurements. All the stability assessment methods performed well for KS_5 , independent of the representation learning method, and showed a higher gradient at the beginning of the experiment which then decreased towards the end of it. Since working with field data, it is also important to note the importance of data acquisition quality which could potentially cause inconsistencies in the results and affect them significantly. It has been proved that the tool wear is the major contributor to

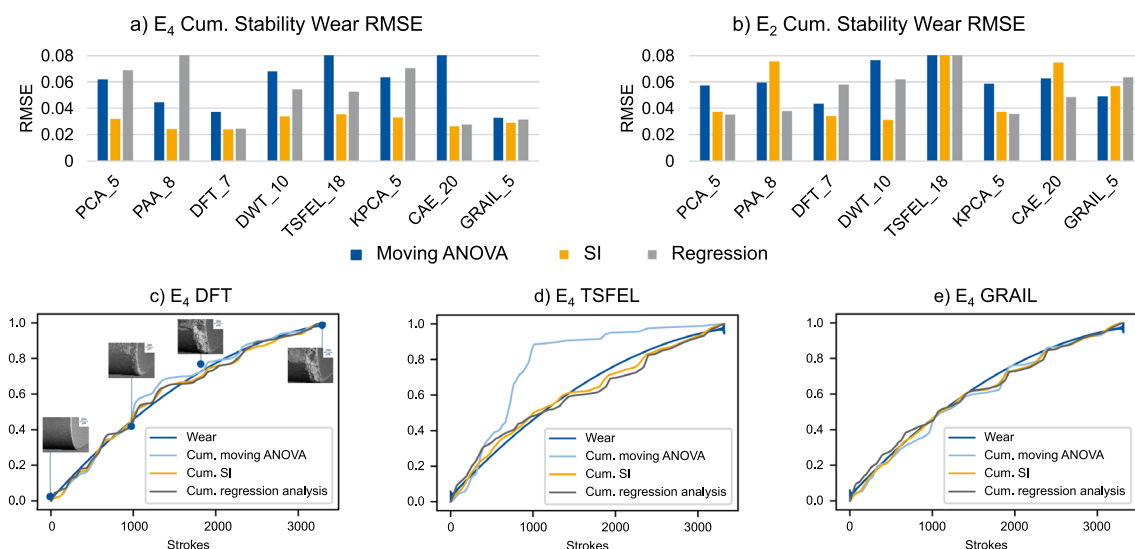


Fig. 8 RMSE of cumulated stability assessment output for **a** E_4 and **b** E_2 , and measured wear plot for **c** DFT, **d** TSFEL, **e** GRAIL representations

the stability of the process, but not the only one. Ramp-up periods, machine stops, or cold starts are events causing the signal instability in clocked sheet metal processing as well.

5 Conclusion

This paper provides an overview of current capabilities of commonly used representation learning methods on time series acquired in clocked sheet metal processing, evaluates them in a theoretical study and explores a practical signal stability assessment as a potential link to the process instability caused by increasing wear of tool components.

The results in Fig. 5 illustrate three things: (i) the dimensions required to appropriately represent variation in measurements, (ii) the trade-off between using different methods to obtain either the global structure or the local structure of a data set, and (iii) that while the study results show a trend, they also show that for some experiments the general trend does not hold. Therefore, more data need to be prepared and collected to validate the degree of freedom required to represent variations in the signal(s) on a larger scale, possibly incorporating similar measurements from other processes. In addition, the study shows that by tracking the fluctuations in the low dimensional representations, an approximation of the wear increase measured in the presented experiment can be achieved.

However, although it has been indicated that changes or variations in representations can be linked to changes in the physical conditions e.g. relating to wear, the complex interplay between various condition changes in a running process environment imposes additional challenges to research the influence of isolated changes on representations and vice

versa. This suggests that an understanding of the so-called process noise [22] or fluctuations in the signals should be better understood. Additionally, the load collective of fine-blanking consists of three process forces that effect each other and the process outcome. Therefore, analysis in the interaction between the forces on the signal side is the next step to investigate the usage of the pipeline presented in this paper.

Acknowledgements Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2023 Internet of Production – 390621612.

Author contributions All authors contributed equally to the work.

Funding Open Access funding enabled and organized by Projekt DEAL.

Availability of data and materials The authors guarantee no restriction of availability of data and code.

Declarations

Conflict of interest The authors have no competing interests.

Ethical approval The authors declare that there are no conflicts with the ethical standards by Springer and the research conducted in this research paper.

Consent to participate A consent was obtained from all individuals included in the study.

Consent to publish The publisher has the consent to the authors to publish the given article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long

as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Klocke F (2013) Manufacturing processes 4: forming. Springer, Berlin Heidelberg, Berlin, Heidelberg
- Ubhayaratne I et al (2017) Audio signal analysis for tool wear monitoring in sheet metal stamping. *Mech Syst Signal Process* 85:809–826. <https://doi.org/10.1016/j.ymssp.2016.09.014>
- Bassiuny AM et al (2007) Fault diagnosis of stamping process based on empirical mode decomposition and learning vector quantization. *Int J Mach Tools Manuf* 47(15):2298–2306. <https://doi.org/10.1016/j.jmachtools.2007.06.006>
- Voss BM et al (2017) Using stamping punch force variation for the identification of changes in lubrication and wear mechanism. *J Phys* 896:012028. <https://doi.org/10.1088/1742-6596/896/1/012028>
- Jin J et al (2000) Diagnostic feature extraction from stamping tonnage signals based on design of experiments. *J Manuf Sci Eng Transact ASME*. <https://doi.org/10.1115/1.538926>
- Hoppe F et al (2019) Feature-based supervision of shear cutting processes on the basis of force measurements: evaluation of feature engineering and feature extraction. *Proc Manuf* 34:847–856. <https://doi.org/10.1016/J.PROMFG.2019.06.164>
- Niemietz P et al (2021) Autoencoder based wear assessment in sheet metal forming. *IOP Conf Ser* 1157:012082. <https://doi.org/10.1088/1757-899X/1157/1/012082>
- Niemietz P et al (2022) Relating wear stages in sheet metal forming based on short- and long-term force signal variations. *J Intell Manuf*. <https://doi.org/10.1007/s10845-022-01979-0>
- Wooluru Y, et al (2015) Approaches for detection of unstable processes: a comparative study. *J Modern Appl Stat Methods* . <https://doi.org/10.22237/masm/1446351360>
- Zhang G et al (2018) Punching process monitoring using wavelet transform based feature extraction and semi-supervised clustering. *Proc Manuf* 26:1204–1212. <https://doi.org/10.1016/j.promfg.2018.07.156>
- Jensen WA et al (2019) Stability assessment with the stability index. *Qual Eng* 31:289–301. <https://doi.org/10.1080/08982112.2018.1497179>
- Agrawal R, et al.: Efficient similarity search in sequence databases, vol. 730 LNCS (1993). https://doi.org/10.1007/3-540-57301-1_5
- Ding H et al (2008) Querying and mining of time series data: experimental comparison of representations and distance measures. *Proc VLDB Endowment*. <https://doi.org/10.14778/1454159.1454226>
- Keogh EJ, Pazzani MJ (2000) A simple dimensionality reduction technique for fast similarity search in large time series databases. In: Terano T, Liu H, Chen ALP (eds) Knowledge discovery and data mining. Current Issues and New Applications: 4th Pacific-Asia Conference, PAKDD 2000 Kyoto, Japan, April 18–20, 2000 Proceedings 4. Springer, Berlin, Heidelberg, pp 122–133. https://doi.org/10.1007/3-540-45571-X_14
- Lopez de Calle K et al (2018) Comparison of automated feature selection and reduction methods on the condition monitoring issue. *Proc Manuf* 16:2–9. <https://doi.org/10.1016/j.promfg.2018.10.150>
- Paparrizos J et al (2019) Grail: efficient time-series representation learning. *Proc VLDB Endowment* 12(11):1762–1777. <https://doi.org/10.14778/3342263.3342648>
- Barandas M et al (2020) Tsfel: time series feature extraction library. *SoftwareX* 11:100456. <https://doi.org/10.1016/j.softx.2020.100456>
- Vankadara LC et al (2018) Measures of distortion for machine learning. *NeurIPS* 2018:31
- Ceruti C et al (2014) Danco: an intrinsic dimensionality estimator exploiting angle and norm concentration. *Pattern Recogn* 47(8):2569–2581. <https://doi.org/10.1016/j.patcog.2014.02.013>
- Ramirez B et al (2006) Quantitative techniques to evaluate process stability. *Qual Eng* 18:53–68. <https://doi.org/10.1080/08982110500403581>
- Behrens BA et al (2016) Advanced wear simulation for bulk metal forming processes. *Numiform*. <https://doi.org/10.15488/1983>
- Liewald M, Bergs T, Groche P, Behrens B-A, Briesenick D, Müller M, Niemietz P, Kubik C, Müller F (2022) Perspectives on data-driven models and its potentials in metal forming and blanking technologies. *Prod Eng*. <https://doi.org/10.1007/s11740-022-01115-0>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.