

54th CIRP Conference on Manufacturing Systems

Dealing with High Dimensional Sequence Data in Manufacturing

Uzma Iffat, Eric Roseren, Mohamed Laib

*Luxembourg Institute for Science and Technology, IT for Innovative Services, Luxembourg** Corresponding author. *E-mail address:* mohamed.laib@list.lu**Abstract**

Advanced manufacturing processes generate large number of parameters and sequences collected from multiple stages of the production pipeline. Consequently, the data dimension explodes due to the high number of features and sequence length. In such a configuration, it is extremely challenging for machine learning algorithms to perform well in limited datasets of few thousand data points. In this paper, we discuss suitable state of the art techniques and propose the application and comparison of few most promising methodologies for effective feature transformation to address this challenge. The results are presented on real-world use cases from the manufacturing sector in Luxembourg.

© 2021 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the scientific committee of the 54th CIRP Conference on Manufacturing System.

Keywords: Industry 4.0; Feature Transformation; Machine Learning; Sequential Data.**1. Introduction**

The use of data analytics and machine learning (ML) in manufacturing has become a very attractive topic in the manufacturing sectors. In recent years, new methodologies are being considered in the literature regarding the use of data-driven approaches for decision-making and in understanding manufacturing processes. The range of applied methods is wide, like for instance, regularised linear regression in predicting product quality, in the semiconductor manufacturing process [1], Extreme learning machine for predicting heat effected zones of the laser cutting process [2]. Other techniques have also been used and investigated, Loyer et al. did a comparison of few machine learning methods to estimate manufacturing cost of jet engine components [3]. A detailed literature review about data analytics in manufacturing is proposed by Sivri and Oztaysi [4]. Cheng et al. emphasised few limitations of the available data mining techniques in dealing with complex and unstructured data issued from manufacturing [5]

Besides, the technology of collecting data is being continuously improved, and many sectors are paying attention to intelligent ways on how to collect data. Consequently, the data size is increasing rapidly. In addition, in manufacturing, the data is recorded from several steps and some parameters are considered as sequences (time series), for instance, the recorded temperature and/or pressure in a specified procedure of the produc-

tion. With the presence of such configuration in data, we need sophisticated methodology to aggregate, structure, and extract useful information from this data in order to explore and understand the manufacturing process, and propose predictive models for improving the quality of products. This research proposes few approaches to deal with this kind of data and handle the problem of high dimensionality and how to aggregate the existing sequences in data. Reducing the dimension of data would help in :

- Increasing the accuracy of learning algorithms.
- Better interpretation of the results provided by learning algorithms.
- Ensure a fast processing of high dimensional data.

This research studies is based on a use case from the hard metal product manufacturing using Powder Metallurgy (PM). An important number of parameters are recorded in the process of powder sintering. During the sintering process, many sequences are monitored such as temperature and pressure. Modelling and simulation of the press and sinter steps have been a strong topic of interest in the PM community. Reviews of the state of the art in computer simulation models performed in [6, 7] explains these limitations as an outcome of lack of experimental data due to expensive testing, inability to account for important environmental factors during the production process and over-simplification of assumptions. With the advent of

advanced data analytics and machine learning, this problem can be redefined using data-driven approaches. A previous work by Konak et. al [8] studies the estimation of shrinkage during hot isostatic pressing (HIP) of nickel based super-alloys for near net shape manufacturing, using feedforward neural network (NN) with two hidden layers. This NN model uses simplified input features such as point statistics from temperature and pressure curves, along with other process attributes. Smith et. al [9] focus on researching AI/ML techniques to solve the inverse problem of recommending suitable input settings to achieve the desirable output in PM manufacturing. Targeted characteristics such as tensile strength, hardness, density and dimension changes are used as inputs for modelling the process setup parameters such as compaction pressure, sintering temperature, sintering time, using a feed forward multilayered neural network. Few other related works such as [10, 11, 12] have studied data driven methods for achieving various other improvements and analysis in the PM manufacturing processes.

This work proposes data-driven approaches to reduce dimensionality and study the stated manufacturing process. We apply three different approaches to deal with the large number of features and sequences. The first one is based on descriptive statistics and some other advanced measures to summarise information from sequences, followed by an unsupervised feature selection to minimise the redundancy among the generated features. The second part focuses on analysing the shapes of the sensor profiles as well as their dissimilarity measures via unsupervised clustering methods. Finally, an Artificial Neural Network (ANN) approach is investigated where a Variational Auto-Encoder(VAE) is used to learn a latent space representation of reduced dimensionality. An encoder-decoder probabilistic network is trained using sequences as input samples and as targets values. A baseline approach using Principal Component Analysis [13] is also included for comparison

The paper is organised as follows. Section 2 describes briefly the problem under study. Section 3 explains the three proposed approaches. In section 4, the results are presented with a comparison of each of the approaches. The comparison is done by using the reduced and transformed feature set from each method in a random forest [14] model setup to predict the shrinkage after sintering. Finally, the conclusion and future improvements are drawn in the last section.

2. Problem Description

As mentioned above, this research considers the case of hard metal product manufacturing using powder metallurgy (PM). The manufacturing process follows the basic press and sinter approach of PM process to form highly precise metal parts and shapes. After the initial steps of pressing and forming, the sintering process is the main step to transform the formed metal powder product to a hard metal product, by cooking the formed powder product in a carefully controlled furnace. During the process of sintering, at very high temperatures, the constituents of the powder mixture melts and fuses to form the hardened

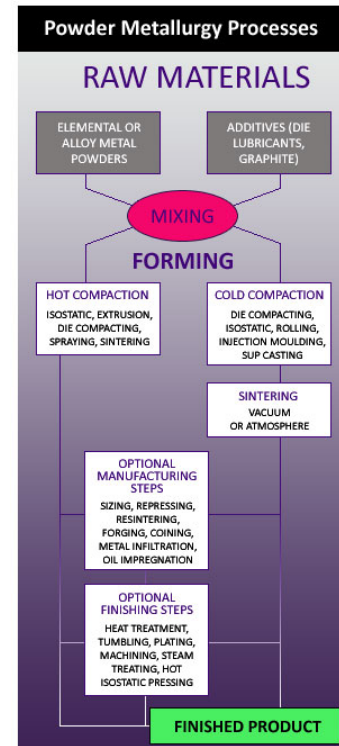


Fig. 1: Overview of the manufacturing process for the use case - work flow in a powder metallurgy manufacturing process by EPMA <https://www.epma.com/powder-metallurgy-process>

product. The furnace parameters are controlled to achieve the desired properties of the target product.

Sintering causes shrinkage in the the dimensions of the product. Sintering conditions are monitored using multiple sensors fitted in the furnace. In this study, 55 furnace sensors are chosen to work with. Each of the sintering load can take variable number of hours depending on the shape and size of the products being sintered. This work focuses on only one type of products, hot rolls for which sintering takes 20 - 24 hrs. The data is recorded at a frequency of 5s by each of the furnace sensors.

The amount of shrinkage which the product undergoes during the sintering process, is a very important variable to control for in the production process. If the product shrinks more that the desired finished dimension, it needs to be discarded with minimal scope of recycling. On the other hand, if the shrinkage is much lesser than expected, the product has to go through the very resource intensive process of grinding for a longer duration than planned. Therefore, it is valuable to predict the shrinkage value based on various attributes of the powder, the initial stages of production and most importantly the sintering phase.

This leads to the main problem statement of this work, how to reduce the dimensionality of multi-dimensional sequences of sensor readings (which are time directional) to utilize the reduced features in the prediction of the shrinkage value.

3. Proposed Approaches

3.1. Principal Component Analysis for Dimension Reduction - Baseline

Principal component analysis (PCA) is commonly used to find latent dimensions of a dataset. It performs an orthogonal linear transformation of the data matrix by finding the eigenvalues and eigenvectors of covariance matrix. The eigenvectors represent the latent dimensions (called PCs) of the data and the magnitude of the eigenvalues can sort the PCs in order of directions which capture the maximum variability of the data. The number of eigenvectors is equal to the rank of the data, which is equal to the number of dimensions in the original data. In order to reduce the number of dimensions in the data, the PCs with relatively lower magnitudes of eigenvalue, thus representing latent dimensions which capture limited variance of the data, can be ignored, effectively allowing dimensionality reduction.

For the furnace sensor data, the data is shaped as $n \times t \times m$ where n denotes the sintering session, t denotes the time steps within the sintering session and m denotes the number of sensors. This data is transformed by stacking the n sintering sessions to get the data shape $(n \times t, m)$. This data includes 3 different sintering furnaces, therefore the data is split by the sintering furnace machine of origin and normalized, before performing PCA. The variance explanation condition is set to 80%, which in this use case reduces the dimensions from 55 in the original to 12 PCs. Further more to achieve a 12 dimensional descriptive feature vector for each furnace run, point statistics such as mean and standard deviation are calculated for each furnace run.

The PCA technique is limited since only linear combinations of data dimensions are considered for the construction of the latent dimensions.

3.2. From feature engineering to redundancy reduction

The first proposed approach generates more feature by summarising the sequences recorded from the furnaces. In fact, several well-known descriptive statistics and advanced measures for time series have been applied such as:

- Mean, variance, skewness, and kurtosis.
- Slopes of temperature increasing (beginning of process) or decreasing (end of process).
- Fisher-Shannon and complexity measures are computed for each sequence [15, 16]. Fisher-Shannon information characterises complexity of non-stationary time series. Furthermore, it offers a better discrimination between sequences, which could help in the construction of predictive models. In fact, Fisher information measure is used to quantify the information in a given sequence and can be computed as follows:

$$I_X = \mathbb{E} \left[-\frac{\partial^2}{\partial x^2} \log f(X) \right] \quad (1)$$

where X is the univariate sequence and $f(X)$ is its probability density function estimated using a kernel density estimation. The Shannon entropy power quantifies the degree of disorder in the sequence and can be obtained:

$$N_X = \frac{1}{2\pi e} e^{2H_X}, \quad (2)$$

where H_X is the differential entropy of X :

$$H_X = \mathbb{E} [-\log f(X)]. \quad (3)$$

These measures have been computed using the R library "FiSh" [17].

- All sequences (data recorded from furnace) are represented by their Intrinsic Dimension (ID) [18]. This measure allows us to get an insight on the amount of information contained in each product sequences. With the presence of redundant features, the ID of data is lower than the space in which they are presented in. In fact, this would indicate if there is a link between the output and the level of redundancy in sequences. The ID was computed using the R library "IDmining" [19].

After computing all these measures, the configuration of the data becomes easier to handle but still with a high number of features. Knowing that these features are correlated, the data contains redundancy that needs to be reduced. In order to minimise this redundancy, we applied an unsupervised feature selection based on the coverage measure [20]. The choice of this method was motivated by the ease of its application. The selected features will be used for the modelling part. Fig. 2 shows the results of the selected features. As expected, the data contains lot of redundant features, and the algorithm selected only 14 features. These selected features will be joined to other categorical features, which was not included in the feature selection process.

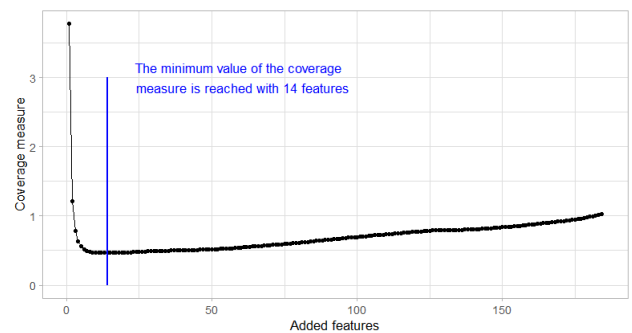


Fig. 2: Unsupervised feature selection applied to dataset containing the new generated features.

3.3. Clustering of sequences

The approach used in section 3.2 focuses on global characteristics computed over the whole time series and relies there-

fore on the appropriate selected measures. However, in some cases, it happens that summary statistics for two widely different curves might be surprisingly similar and in order to differentiate them another approach might be required.

In this section we focused on comparing time series based on the profile of their sensor curves. The idea is to discriminate profiles that are deviating from a cluster centroid (e.g the average profile of a cluster) also called time series prototype. We essentially performed clustering on multivariate time series using a dissimilarity distance, a prototype function to calculate each cluster centroid and finally a partitioning algorithm. Because most sensor signals were exhibiting significant shifting and lagging we proceeded with Dynamic Time Warping (DTW) as our dissimilarity measure. Regarding centroids, there exists several methods to compute time series prototypes such as simply computing the arithmetic mean or median of all series within a cluster. More recent techniques include Dynamic Time Warping Barycenter Averaging (DBA) [21], Partition Around Medoid (PAM) and shape extraction, which is part of the k-Shape algorithm described in [22]. We tested several prototype functions and decided to proceed with DBA. With regards to the partitioning clustering task, we started by grouping each sensor profiles to a pre-defined number of cluster. This number was chosen after several test runs to check how clusters were formed and if the time series could efficiently be discriminated. Figure 3 displays an example of clusters computed for a type of sensor. Our final choice of clustering algorithm was a partitioning algorithm using dynamic time warping (DTW) as dissimilarity measure which was then coupled to a DTW barycenter averaging (DBA) prototype function. More information regarding time series clustering using this method can be found in [23, 24].

Once the clustering part of the profiles done, the DTW distance between each series in the data and its corresponding centroid was then computed. The larger the dissimilarity measure for a profile, the further apart this process is compared to other profiles. Ideally this will create a set of features with reduced number of dimensions that encapsulate most of the variation happening in the manufacturing process.

3.4. Variation Autoencoders

Autoencoders (AE) [25] are unsupervised representational learning techniques which impose a bottleneck using neural network architecture to force a dimension-reduced knowledge representation. The network learns by minimizing the reconstruction loss between the original input and reconstructed input created from the limited information which can flow through the imposed bottleneck. While PCA can learn linear hyperplanes, AEs are capable of learning nonlinear manifolds.

Variational autoencoders (VAE) [26] is the probabilistic variant of AE which is regularized to avoid overfitting and the dimension reduced latent space created by the VAE has properties favourable to a generative process such as continuity and completeness.

In the given use case, the furnace data for each sintering session can be arranged as a 2D matrix with time as one dimension

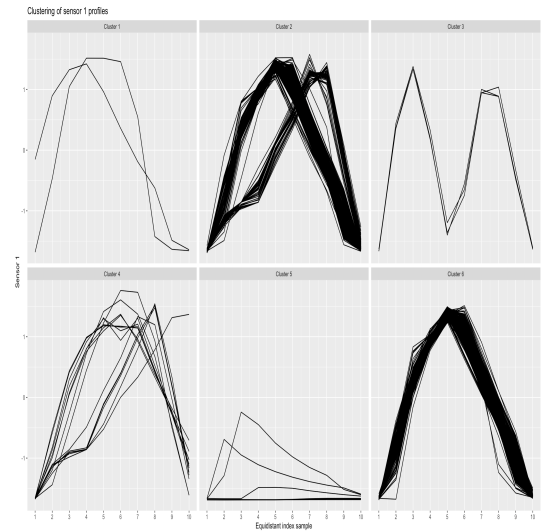


Fig. 3: Partitional clustering using DTW distance and DBA prototype function.

sion, in which directionality matters and sensors as the other dimension, with no directionality. The time length of sintering can be variable, so to obtained fixed dimension inputs, the sequence data is clipped at the n th time step, and only sequences of length equal or greater than n are considered.

Data augmentation techniques can also be applied to the sequence data to control overfitting. [27] surveys a range of techniques for time series data. We propose an interleaved sampling technique for data augmentation. If the sequence data collected from the sensors is sampled at a high frequency, the augmented sequences can be created by down-sampling from the original sequence. As depicted in Fig 3(a), if $t_1, t_2, t_3, \dots, t_n$ are the time steps in the original sequence, the augmented sequences will have time steps $[t_i, t_{i+k}, t_{i+2k}, \dots]$ where i is the time-shifted starting points such as $t = 1, 2, \dots, n$ and k is the down-sampling rate. The resulting augmented sintering session can be visualised as Fig 4.

The encoder-decoder architecture in the VAE model uses 1-D convolutional blocks for convolving in the time dimension. The model is trained for 100 epochs and loss minimized is a weighted sum of reconstruction loss, composed of binary cross-entropy loss for the 2D input array and a KL divergence loss to regularize the latent data distribution.

The number of latent dimensions can be chosen empirically depending on the final task such as performance of a prediction model which uses the latent embeddings. For the ease of visualization, the model is constructed for 2 dimensions and the resulting manifold can be as shown in Fig 5.

A visualisation of the reconstructed input array from a test sample clearly shows the the VAE encodings are able to capture the high level details of the furnace sequence arrays. Using higher latent space dimension, reduces the reconstruction loss. See Fig 6.

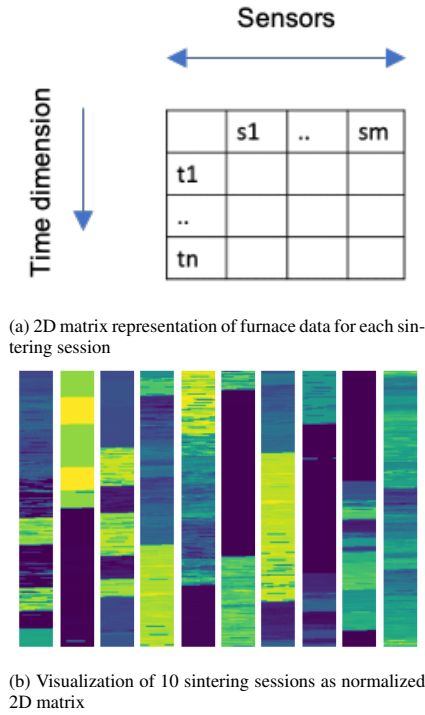


Fig. 4: Input Data for VAE

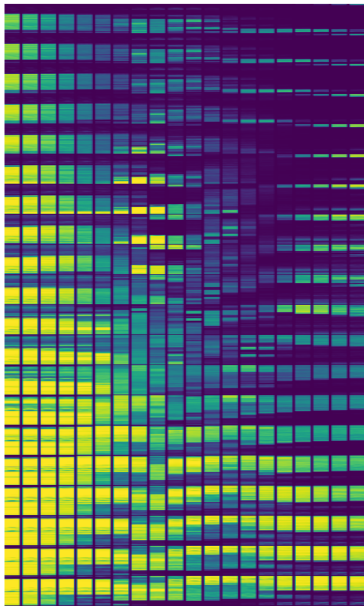


Fig. 5: Grid of furnace arrays sampled from the two dimensional latent space

4. Results and discussions

As mentioned above, each approach provided new features that are useful in the shrinkage prediction model. A comparison of the effectiveness of the dimension reduction methods has been drawn by using the features generated by each method in the prediction model which uses the random forest algorithm. The train and test sample has been maintained constant between

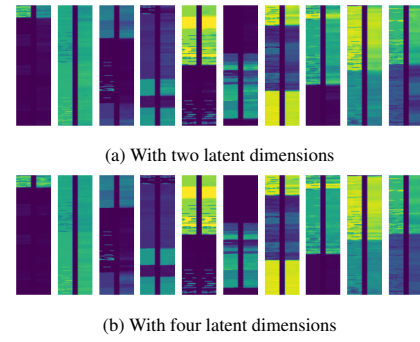


Fig. 6: Pairs of original and reconstructed furnace arrays from test sample

each experiment. To summarise, random forest has been applied on features defined by the following approaches:

- Feature Engineering and Dimension Reduction (**FEDR**), as described in section 3.2.
- Clustering Sequences (**CS**), explained in section 3.3.
- Variational AutoEncoder approach (**VAE**), from section 3.1.

An additional well-known approach based on principal component analysis is used as baseline for this comparison (**PCs**).

Table 1 shows the results of the prediction model applied on each dataset. The results are quite similar with a slight advantage for VAE approach. Having the least mean squared error but relatively higher mean absolute error implies that VAE methods has lead to lesser instances of large errors.

The error performance advantage of VAE is justified due to the fact that VAE features capture all available information of the sequences, while FEDR and CS approaches summarise each sequences in few measures. In addition, VAE performs better than the baseline PCA approach since VAE can detect even non-linear relationship when performing the feature extraction.

Although both PCA and VAE create new features from combinations of original features, the VAE encoder is less interpretable as compared to the PCA. The FEDR and CS approach are more explainable in terms of original feature contributions.

Another advantage of using the VAE approach is when the predictive model also used to suggest initial furnace setup parameters. The generative nature of the VAE allows the sampling of original sensor sequences from the reduced latent space.

Table 1: Regression results using a Random Forest algorithm (mean of 20 runs and the standard deviation).

	PCs	FEDR	CS	VAE
MSE	0.264(0.004)	0.253(0.001)	0.284(0.005)	0.249 (0.01)
MAE	0.363(0.004)	0.352 (0.009)	0.379(0.004)	0.359(0.012)

5. Conclusion

The research studied a challenging case study of hard metal manufacturing. The data contains important number of features recorded during the production process. Furthermore, some features were represented by temporal sequences recorded when the product is in the furnace (e.g. temperature, pressure). The paper showed few methodologies to deal with such configuration of data and reduce the dimensionality.

The proposed approaches can be seen from two important perspectives. The first one is the capability to interpret the model results and to investigate which features are important (e.g. FEDR approach). The second one is the ability of getting the best prediction (e.g. VAE approach). The choice of which approach to use is motivated by the objective of the study. Further research could be in performing a multivariate time series clustering to improve the results of CS approach. In addition, using more informative time series measures to summarise better the sequences from the furnace could improve the prediction accuracy, by suggesting other relevant features.

Acknowledgements

The present work is partly supported by the Luxembourg National Research Fund (BRIDGES18/IS/13307925), and the Ministry of Economy in Luxembourg under the RDI Law.

References

- [1] M. Melhem, B. Ananou, M. Ouladsine, J. Pinaton, Regression methods for predicting the product quality in the semiconductor manufacturing process, IFAC-PapersOnLine 49 (12) (2016) 83 – 88, 8th IFAC Conference on Manufacturing Modelling, Management and Control MIM 2016. doi:<https://doi.org/10.1016/j.ifacol.2016.07.554>.
- [2] O. Anicic, S. Jović, H. Skrijelj, B. Nedić, Prediction of laser cutting heat affected zone by extreme learning machine, Optics and Lasers in Engineering 88 (2017) 1 – 4. doi:<https://doi.org/10.1016/j.optlaseng.2016.07.005>.
- [3] J.-L. Loyer, E. Henriques, M. Fontul, S. Wiseall, Comparison of machine learning methods applied to the estimation of manufacturing cost of jet engine components, International Journal of Production Economics 178 (2016) 109 – 119. doi:<https://doi.org/10.1016/j.ijpe.2016.05.006>.
- [4] M. Sami Sivri, B. Oztaysi, Data Analytics in Manufacturing, Springer International Publishing, Cham, 2018, pp. 155 – 172. doi:https://doi.org/10.1007/978-3-319-57870-5_9.
- [5] Y. Cheng, K. Chen, H. Sun, Y. Zhang, F. Tao, Data and knowledge mining with big data towards smart production, Journal of Industrial Information Integration 9 (2018) 1 – 13. doi:<https://doi.org/10.1016/j.jii.2017.08.001>.
- [6] S.-H. Chung, H. Company, y.-s. Kwon, I. Cetatech, J. Seong, S. J. Park, Modeling and Simulation of Press and Sinter Powder Metallurgy, ASM Handbook 22 (Jan. 2010).
- [7] R. M. German, Computer modeling of sintering processes, International Journal of Powder Metallurgy 38 (2002) 48–66.
- [8] A. Konak, S. Kulturel-Konak, A. E. Smith, I. Nettleship, Estimation of shrinkage for near net-shape using a neural network approach, Journal of Intelligent Manufacturing 14 (2) (2003) 219–228. doi:<https://doi.org/10.1023/A:1022907615088>.
- [9] L. N. Smith, R. M. German, M. L. Smith, A neural network approach for solution of the inverse problem for selection of powder metallurgy materials, Journal of Materials Processing Technology 120 (1) (2002) 419–425. doi:[https://doi.org/10.1016/S0924-0136\(01\)01198-0](https://doi.org/10.1016/S0924-0136(01)01198-0).
- [10] R. P. Cherian, L. N. Smith, P. S. Midha, A neural network approach for selection of powder metallurgy materials and process parameters, Artificial Intelligence in Engineering 14 (1) (2000) 39–44. doi:[https://doi.org/10.1016/S0954-1810\(99\)00026-6](https://doi.org/10.1016/S0954-1810(99)00026-6).
- [11] H. Khorsand, M. Arjomandi, H. Abdoos, S. H. Sadati, Application of Artificial Neural Network for Prediction of Heat Treated Sintered Steels Properties (2008). doi:<https://doi.org/10.4028/www.scientific.net/DDF.273-276.323>.
- [12] P. Radha, G. Chandrasekaran, N. Selvakumar, Generalized Neural Network Model to Predict the Properties of Sintered Al - Fe Composite, in: International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007), Vol. 1, 2007, pp. 290–296, ISSN: null. doi:<https://doi.org/10.1109/ICCIMA.2007.285>.
- [13] J. Shlens, A tutorial on principal component analysis, arXiv preprint arXiv:1404.1100 (2014).
- [14] L. Breiman, Random forests, Machine Learning 45 (2001) 5 – 32. doi:<https://doi.org/10.1023/A:1010933404324>.
- [15] F. Guignard, M. Laib, F. Amato, M. Kanevski, Advanced analysis of temporal data using fisher-shannon information: Theoretical development and application in geosciences, Frontiers in Earth Science 8 (2020) 255. doi:<https://doi.org/10.3389/feart.2020.00255>. URL <https://www.frontiersin.org/article/10.3389/feart.2020.00255>
- [16] F. Amato, M. Laib, F. Guignard, M. Kanevski, Analysis of air pollution time series using complexity-invariant distance and information measures, Physica A: Statistical Mechanics and its Applications 547 (2020) 124391. doi:<https://doi.org/10.1016/j.physa.2020.124391>. URL <http://www.sciencedirect.com/science/article/pii/S0378437120301497>
- [17] F. Guignard, M. Laib, FiSh: Fisher-Shannon Method, R package version 1.0 (2019). URL <https://CRAN.R-project.org/package=FiSh>
- [18] J. Golay, M. Kanevski, A new estimator of intrinsic dimension based on the multipoint Morisita index, Pattern Recognition 48 (12) (2015) 4070 – 4081. doi:<https://doi.org/10.1016/j.patcog.2015.06.010>.
- [19] J. Golay, M. Laib, IDmining: Intrinsic Dimension for Data Mining, R package version 1.0.6 (2019). URL <https://CRAN.R-project.org/package=IDmining>
- [20] M. Laib, M. Kanevski, A new algorithm for redundancy minimisation in geo-environmental data, Computers & Geosciences 133 (2019) 104328.
- [21] F. Petitjean, A. Ketterlin, P. Gançarski, A global averaging method for dynamic time warping, with applications to clustering, Pattern Recognition 44 (3) (2011) 678–693.
- [22] J. Paparrizos, L. Gravano, k-shape: Efficient and accurate clustering of time series, in: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, 2015, pp. 1855–1870.
- [23] A. Sardá-Espinosa, Time-series clustering in R using the dtwclust package., R J. 11 (1) (2019) 22.
- [24] A. Sardá-Espinosa, dtwclust: Time Series Clustering Along with Optimizations for the Dynamic Time Warping Distance, R package version 5.5.6 (2019). URL <https://CRAN.R-project.org/package=dtwclust>
- [25] P. Baldi, Autoencoders, unsupervised learning, and deep architectures, in: Proceedings of ICML workshop on unsupervised and transfer learning, 2012, pp. 37–49.
- [26] D. P. Kingma, M. Welling, Auto-encoding variational bayes, arXiv preprint arXiv:1312.6114 (2013).
- [27] B. K. Iwana, S. Uchida, An empirical survey of data augmentation for time series classification with neural networks, arXiv preprint arXiv:2007.15951 (2020).