

Fault Diagnostics by Conceptors-Aided Clustering

Mingjing Xu

Energy Department, Politecnico di Milano, Italy. E-mail: mingjing.xu@polimi.it

Piero Baraldi*

Energy Department, Politecnico di Milano, Italy. E-mail: piero.baraldi@polimi.it

Enrico Zio

Energy Department, Politecnico di Milano, Italy. E-mail: enrico.zio@polimi.it

MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France.

Eminent Scholar, Department of Nuclear Engineering, College of Engineering, Kyung Hee University, Republic of Korea.

Aramis Srl, Milano, Italy.

Fault diagnostics in practice faces the challenge of dealing with unlabelled time series that have long-term temporal dependencies. Inspired by the idea of representing temporal patterns by a mechanism of neurodynamical pattern learning, called Conceptors, we propose an unsupervised clustering method for identifying the degradation state of industrial equipment. Conceptors are used to represent the dynamic behaviour of the degradation trajectories and spectral clustering is used to group the Conceptors in homogenous classes of similar degradation states. The proposed method is applied to a case study of literature. The results show that the accuracy of the fault diagnosis is satisfactory.

Keywords: Fault diagnostics, time series clustering, Conceptors, reservoir computing.

1. Introduction

The objective of this work is to develop a method for diagnosing the degradation state of working equipment using signals measured during operation and assuming that only a few unlabelled run-to-failure degradation trajectories are available, as is the case in many practical applications.

In the proposed method, each complete run-to-failure degradation trajectory is unsupervisedly clustered into several non-overlapped time-series segments representing different degradation states.

Unsupervised learning is an important topic in machine learning for time series segmentation (Saravanan and Ramachandran 2010) and pattern recognition (Zheng et al. 2018). In practical applications, it is strongly related to fault diagnostics. Indeed, it can provide abstract representations of the raw data measured from equipment, for distinguishing healthy and faulty conditions (Sameer Al-Dahidi et al. 2018; Baraldi et al. 2015; Liu et al. 2018; Martins, Pires, and Pires 2007). In (Sameer Al-Dahidi et al. 2018), a framework for reconciliating data clusters obtained by unsupervised consensus

clustering from a fleet of nuclear power plants turbines is proposed. In (Baraldi et al. 2015), a diagnostic methodology based on unsupervised spectral clustering combined with fuzzy C-means (*FCM*) is developed for identifying groups of similar shut-down transients of nuclear turbines. In (Liu et al. 2018), a Categorical Adversarial Autoencoder (*CatAAE*) is proposed for unsupervised learning for fault diagnostics of rolling bearings.

Conceptor is a mechanism of neurodynamical temporal pattern representation proposed in (Jaeger 2016). Considering a reservoir, i.e. a randomly generated and sparsely connected Recurrent Neural Network (*RNN*) (Lukouevius and Jaeger 2009), Conceptors characterize the geometries of the temporal states of the reservoir neurons in the form of square matrices (Jaeger 2016), achieving a direction-selective damping of high-dimensional reservoir states (Qian and Zhang 2017).

A case study taken from literature (Qiu et al. 2006) is used to verify the performance of the method proposed in this work and to compare it with an unsupervised diagnostics technique based on the use of Dynamic Time Warping

(DTW).

The remaining of the paper is organized as follows: Section 2 states the problem and illustrates the objectives of the work; Section 3 introduces the background of the proposed method and Section 4 specializes the proposed method to equipment fault diagnostics; Section 5 describes the performance metrics used to evaluate the proposed method; Section 6 introduces the case study and discusses the obtained results; finally, some conclusions and remarks are drawn in Section 7.

2. Problem Statement

We consider a practical situation in which S signals have been measured during R run-to-failure lives of a population of similar equipment. The generic r -th ($r = 1, \dots, R$) life trajectory is represented by a matrix \mathbf{P}^r , whose generic component $p_{s,i}^r$ represents the value of signal s taken at time t_i , $s = 1, \dots, S$ and $i = 1, \dots, N_r$, where N_r is the number of measurements taken during the life of equipment r .

The objective is to estimate the degradation state d_i of a piece of equipment at a generic time t_i , $i = 1, \dots, N_r$, given the measurements $p_{1:S,1:N_r}$ of S signals collected until time t_i . We assume that the degradation state can be represented using a discrete set of $g \in \mathbb{N}^+$ levels, assumed to be the same for all equipment among the population.

3. Background

3.1 Reservoir Computing

A reservoir is a large, randomly connected recurrent neural network that receives in input time-varying signals (Lukoueviius and Jaeger 2009). In this work, we consider Echo State Networks (ESNs) (Ferreira, Ludermir, and De Aquino 2013), whose architecture is characterized by S input neurons and a reservoir with $N \gg 1$ internal neurons (Fig. 1). The matrix \mathbf{W}_{in} of size $N \times S$ contains the weights of the connections from the input neurons to the internal neurons. The matrix \mathbf{W} of size $N \times N$ contains the weights of the connections among the internal neurons. The elements of \mathbf{W}_{in} and \mathbf{W} are randomly initialized from a uniform distribution $U(-1,1)$ and remain unchanged during learning (Xu et al. 2020a; 2020b). A crucial, enabling precondition for the ESN

learning algorithms to function is the *Echo State Property (ESP)*, which is that the effect of the current states of the reservoir internal neurons and of the input on a future state should vanish gradually as time passes, and not get amplified or even persist (Ferreira, Ludermir, and De Aquino 2013). A sufficient condition for ESP in standard ESNs is $\rho(|\mathbf{W}|) < 1$, where $\rho(|\mathbf{W}|)$ is the spectral radius of $|\mathbf{W}|$: i.e. the magnitude of the largest eigenvalue of the absolute of \mathbf{W} (Ferreira, Ludermir, and De Aquino 2013).

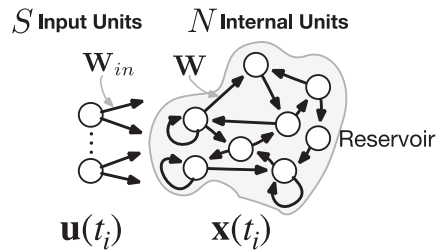


Fig. 1. ESN architecture.

The reservoir state at time t_i is updated according to:

$$\mathbf{x}(t_i) = \mathbf{f}(\mathbf{W}_{in} \mathbf{u}(t_i) + \mathbf{W} \mathbf{x}(t_{i-1})) \quad (1)$$

where $\mathbf{u}(t_i)$ denotes the input at time t_i , i.e. the input measurements, and $\mathbf{x}(t_i) \in \mathbb{R}^N$ represents the reservoir states. The state activation function $\mathbf{f} = (f_1, \dots, f_N)^T$ is a sigmoid of function, such as \tanh , applied component-wise.

3.2 Conceptors

Conceptors can be described as filters characterizing the temporal reservoir activation patterns (Jaeger 2016). The Conceptor matrix \mathbf{C} is defined as the linear transformation of the reservoir state $\mathbf{x}(t_i)$ that minimizes the loss function:

$$\mathbb{E}_{t_1=t_1, t_2, \dots, t_l} [\|\mathbf{x}(t_i) - \mathbf{C} \mathbf{x}(t_i)\|^2] + \alpha^{-2} \|\mathbf{C}\|_{Fro}^2 \quad (2)$$

where α is a control parameter called *aperture*, $\|\cdot\|_{Fro}$ is the Frobenius norm (Jaeger 2016) and l is the length of the time sequence. The closed-form solution of the optimization problem is:

$$\mathbf{C} = \mathbf{R}(\mathbf{R} + \alpha^{-2} \mathbf{I})^{-1} \quad (3)$$

where $\mathbf{R} = \mathbb{E}_{t_i=t_1, t_2, \dots, t_l}[\mathbf{x}(t_i)\mathbf{x}(t_i)^T]$ is the $N \times N$ correlation matrix of $\mathbf{x}(t_i)$ and \mathbf{I} is the $N \times N$ identity matrix. In intuitive terms, \mathbf{C} is a soft projection matrix on the linear subspace where the samples of $\mathbf{x}(t_i)$ lie.

3.3 Spectral Clustering

Spectral clustering is based on the construction of a similarity graph $G = (V, E)$ where $V = \{v_1, \dots, v_l\}$ identifies the set of vertices and $E = \{v_1 - v_2, v_1 - v_3, \dots, v_l - v_{l-1}\}$ the set of edges connecting the vertices. Each vertex represents an object, and the weight associated to the edge connecting two generic vertices v_m and v_n is the measure of similarity between objects m and n , denoted by A_{mn} , $m, n = 1, 2, \dots, l$. Spectral clustering aims at defining partitions of vertexes, which will be referred to as clusters, such that the edges between objects belonging to the same partition have associated large weights (objects within the same cluster are similar), whereas the edges between objects belonging to different partitions have associated small weights (objects in different clusters are dissimilar). The spectral clustering technique entails five main steps (Di Maio et al. 2013):

1. Build the normalized Laplacian matrix \mathbf{L} . This requires the computation of the diagonal degree matrix, \mathbf{A} , whose entries A_1, A_2, \dots, A_l are: $A_m = \sum_{n=1}^l A_{mn}$, from which the normalized Laplacian matrix is obtained:

$$\mathbf{L} = \mathbf{I} - \mathbf{A}^{-\frac{1}{2}} \mathbf{A} \mathbf{A}^{-\frac{1}{2}} \quad (4)$$

where \mathbf{I} is the identity matrix of size $l \times l$.

2. Compute the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_l$ and corresponding eigenvectors b_1, b_2, \dots, b_l of the normalized Laplacian matrix \mathbf{L} , in which the eigenvalues are sorted from smallest to largest. Select the first k smallest eigenvalues $\lambda_1, \dots, \lambda_k$ and corresponding eigenvectors b_1, \dots, b_k (Von Luxburg 2007).
3. Build the matrix \mathbf{B} of size $l \times k$, in which the k columns are the eigenvectors b_1, \dots, b_k found in step 2. A matrix \mathbf{Z} is, then, obtained by normalizing \mathbf{B} rows (Von Luxburg 2007):

$$z_{m\tilde{n}} = b_{m\tilde{n}} / \left(\sum_{\tilde{n}=1}^k b_{m\tilde{n}}^2 \right)^{1/2}, \quad m = 1, \dots, l, \tilde{n} = 1, \dots, k \quad (5)$$

It has been shown that this change of data

representation enhances the cluster properties in the data, i.e., it allows more easily identifying clusters (Von Luxburg 2007).

4. Apply a clustering algorithm to the rows of the matrix \mathbf{Z} , each one representing an object in the space of the first k normalized eigenvectors. In this work, we use the k -means clustering algorithm with k clusters Π_1, \dots, Π_k (Uw et al. 2001).

4. Proposed method for fault diagnostics

The method proposed for equipment fault diagnostics is based on the two phases of generating Conceptors to represent the degradation trajectories and spectral-clustering them.

4.1.1 Generating Conceptors

The objective of the first phase is to represent the generic r -th degradation trajectory, $p_{1:s,1:i}^r \in \mathbb{R}^{S \times i}$, from time t_1 to the generic time t_i , by means of the Conceptor $\mathbf{C}_i^r \in \mathbb{R}^{N \times N}$.

To this aim, we first fix the reservoir architecture by setting a spectral radius $\rho(|\mathbf{W}|)$ which satisfies $\rho(|\mathbf{W}|) < 1$ for ensuring the echo state property (Yildiz, Jaeger, and Kiebel 2012), the number of reservoir internal neurons N (Qiao et al. 2016) and the connectivity c (Büsing, Schrauwen, and Legenstein 2010), i.e. the ratio between the number of connections in the reservoir and the number N^2 of all possible connections, which is taken inversely proportional to the reservoir size, i.e., $c = M/N$, $M \in \mathbb{R}^+$, for performance improvement and computational complexity reduction (Qiao et al. 2016; Lukoševičius 2012).

TABLE 1 reports the settings of the Conceptor hyper-parameters used in this work. Since larger the reservoir size N , larger the Memory Capacity (MC) which quantifies the memory span of the ESN (Qiao et al. 2016; Büsing, Schrauwen, and Legenstein 2010), but also, larger the computational burden, a trade-off value of $N = 100$ is chosen for the reservoir size. The spectral radius $\rho(|\mathbf{W}|)$ is set equal to 0.9 to allow longer retainment of the past system states, which requires a large value, while ensuring the echo state property, which requires $\rho(|\mathbf{W}|) \leq 1$ (Yildiz, Jaeger, and Kiebel 2012). The connectivity c is set equal to $5/N$ to guarantee

proper *MC* without extensively increasing the computational burden (Qiao et al. 2016). Conceptor aperture α , which can be interpreted as the scaling factor of the reservoir state in Eq. (3), is set equal to 1 according to (Jaeger 2016).

The Conceptor matrix \mathbf{C}_i^r is computed by applying Eqs. (1) and (3) to the multi-variate time series $p_{1:S,1:i}^r$, with the reservoir states at time t_0 , $\mathbf{x}(t_0)$, initialized to zeros to avoid unnecessary uncertainty. The correlation matrix \mathbf{R} is computed by adopting an iterative updating procedure:

$$\mathbf{R}^{(i)} = \mathbf{R}^{(i-1)} \cdot \frac{i-1}{i} + \mathbf{x}(t_i)\mathbf{x}(t_i)^T \cdot \frac{1}{i} \quad (6)$$

where $\mathbf{R}^{(i)}$ denotes the correlation matrix at time t_i and $\mathbf{R}^{(0)}$ the null matrix.

TABLE 1. Setting of the hyper-parameters of the proposed method

Hyper-parameters	Values
Reservoir size (N)	100
Spectral radius ($\rho(\mathbf{W})$)	0.9
Connectivity (c) \times Reservoir size (N)	5
Conceptors aperture (α)	1
Similarity measure parameter (σ)	1
Maximum possible number of clusters (K)	8

4.1.2 Clustering

Let $\{\mathbf{C}_i^r\}_{i=1,\dots,N_r}$ be the set of Conceptor matrixes, obtained from the generic r -th trajectory. The objective of this module is to cluster them into $\hat{g} > 1$ clusters $\Pi_1^r, \dots, \Pi_{\hat{g}}^r$, where \hat{g} is the estimation of the number g of degradation states.

Since traditional clustering methods cannot be directly used for clustering matrixes, and the vectorization of a matrix causes an exponential increase of the problem dimensionality, we resort to the spectral clustering algorithm (Von Luxburg 2007).

The distance metric between two Conceptors \mathbf{C}_i^r and \mathbf{C}_j^r , $i, j = 1, \dots, N_r$, is computed using the Frobenius norm of the difference matrix of the two Conceptors (Jaeger 2016):

$$D_{ij}^r = \|\mathbf{C}_i^r - \mathbf{C}_j^r\|_{fro} \quad (7)$$

Then, the similarity measure A_{ij}^r between Conceptors \mathbf{C}_i^r and \mathbf{C}_j^r taken from the r -th training trajectory is (Von Luxburg 2007):

$$A_{ij}^r = \exp\left(-\frac{(D_{ij}^r)^2}{2\sigma^2}\right) \quad (8)$$

where σ is a parameter which can be set by the analyst to shape the desired interpretation of similarity: the larger the value of σ , the wider the width of the neighborhoods and the slower the decline of similarity with respect to the Conceptors distance.

Finally, the spectral clustering algorithm of Section 3.3 is applied to obtain the k clusters of Conceptors $\{\Pi_1^r, \dots, \Pi_k^r\}^{(k)}$ of the r -th trajectory. The optimal number of clusters k_r^* is set by maximizing the Silhouette Index SI_k^r (S. Al-Dahidi et al. 2015; Rousseeuw 1987) (in Appendix), which is a measure of how much an object is similar to another object of its own cluster compared to an object of other clusters:

$$k_r^* = \underset{k=2,\dots,K}{\operatorname{argmax}} SI_k^r \quad (9)$$

where K is the maximum possible number of degradation states, which is set equal to 8 (Baraldi et al. 2015).

The procedure is repeated for all the training trajectories. The estimated number of degradation states \hat{g} to be considered for clustering all degradation trajectories is taken equal to the value appearing most often in the set $\{k_r^*\}_{r=1,\dots,R}$. Then, for each run-to-failure trajectory, the clusters identified by the spectral clustering algorithm when \hat{g} clusters were searched are used.

To label the Conceptors clusters $\Pi_1^r, \dots, \Pi_{\hat{g}}^r$, ($r = 1, \dots, R$) as specific degradation states, we observe that the degradation process is typically monotonic, i.e. $d_i^r \leq d_{i+1}^r$, and continuous, i.e. in one time step the component degradation can upmost increase of one level, if i.e. $\mathbf{C}_i^r \in \Pi_{d_i^r}^r \rightarrow \mathbf{C}_{i+1}^r \in [\Pi_{d_i^r}^r, \Pi_{d_i^r+1}^r]$.

5. Performance Metrics

Let us consider the r -th generic run-to-failure trajectory $p_{1:S,1:N_r}^r$. For the sake of simplicity of the notation, m is used to indicate the segment of the trajectory $p_{1:S,1:m}^r$ going from time t_1 to time t_m . The *Accuracy* metric assessing the performance of the clustering of the r -th trajectory is:

$$Accuracy_r = \frac{\sum_{m=1}^{N_r} \delta(y_m, c_m)}{N_r} \quad (10)$$

where y_m and c_m are the ground-truth and the assigned equipment degradation level at time t_m ,

respectively, $\delta(y_m, c_m)$ is the delta function, which is equal to 1 if $y_m = c_m$ and 0, otherwise. The *Accuracy* metric is in the range [0, 1], with 1 indicating the most satisfactory performance.

The Mutual Information (*MI*) metric measures the degree of dependency between the ground-truth and assigned degradation states (Cakir et al. 2019). It is preferred to the *MI* metric when the dataset is imbalanced (Jain 2010); the *MI* metric assessing the performance in the clustering of the r -th trajectory is defined by (Ye et al. 2018):

$$MI_r = \sum_{\hat{\gamma}=1}^{\hat{g}} \sum_{\beta=1}^g p_r(\beta, \hat{\gamma}) \log \frac{p_r(\beta, \hat{\gamma})}{p_r(\beta)p_r(\hat{\gamma})} \quad (11)$$

where, considering a randomly sampled segment m extracted from the r -th training trajectory, $p_r(\beta)$ and $p_r(\hat{\gamma})$ are the probabilities that its true and assigned degradation states are β and $\hat{\gamma}$, respectively, $p_r(\beta, \hat{\gamma})$ is the joint probability that its true degradation state is β and its assigned degradation state is $\hat{\gamma}$ at the same time, and g and \hat{g} are the numbers of true and assigned degradation states, respectively. *NMI* normalizes *MI* in the range [0,1] (Ye et al. 2018):

$$NMI_r = \frac{MI_r}{(H[y_1, y_2, \dots, y_{N_r}] + H[c_1, c_2, \dots, c_{N_r}]) / 2} \quad (12)$$

where $H[y_1, y_2, \dots, y_{N_r}]$ and $H[c_1, c_2, \dots, c_{N_r}]$ denote the entropy of the true and assigned degradation states obtained from the clustering of the r -th trajectory, respectively:

$$H[y_1, y_2, \dots, y_{N_r}] = - \sum_{\beta=1}^g p_r(\beta) \log p_r(\beta),$$

$$H[c_1, c_2, \dots, c_{N_r}] = - \sum_{\hat{\gamma}=1}^{\hat{g}} p_r(\hat{\gamma}) \log p_r(\hat{\gamma}) \quad (13)$$

A *NMI* value of 1 indicates the most satisfactory performance.

6. Case Study

The bearing dataset provided by the Center for Intelligent Maintenance Systems (IMS) of the, University of Cincinnati (Qiu et al. 2006) is considered. A shaft is coupled to an AC motor

and is rotating at 2000 rpm while a 6000 lbs load is installed on it. Four force-lubricated bearings are mounted on the shaft and two accelerometers are placed close to each bearing for measuring the vibrations at a sampling frequency of 20 kHz. The bearing test rig and the position of the sensors are shown in Fig. 2. Time windows of length equal to 1 second, containing 20480 acceleration values, are measured every 10 min. The degradation trajectories of bearings No.3 and 4, which are the only ones running to failure in experiment 1 (Qiu et al. 2006), are considered. Each trajectory is formed by $N_r = 2156$, ($r = 1, 2$), 1-second length time windows, where $r = 1, 2$ denote bearing No.3 and 4, respectively. The literature works that have considered the same case study have detected the onset of abnormal conditions on bearing No.3 at times 1617 (Qiu et al. 2006), 2027 [68] and on bearing No.4 at times 1617 (Qiu et al. 2006), 1760 (Yu 2012) and 1641 [68]. This work, therefore, assumes that a transition from the healthy state to the degraded state has occurred at the average times among those reported in literatures, which are 1822 for bearing No. 3 and 1672 for bearing No.4, and we refer to these two quantities as reference transition times.

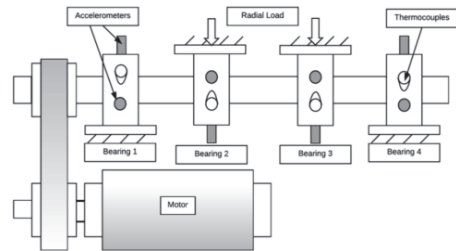


Fig. 2. Bearing test rig (Qiu et al. 2006).

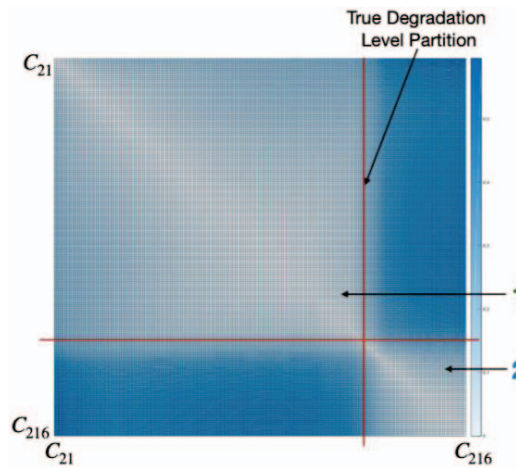


Fig. 3. Pairwise Conceptors distance matrix on bearing No.4.

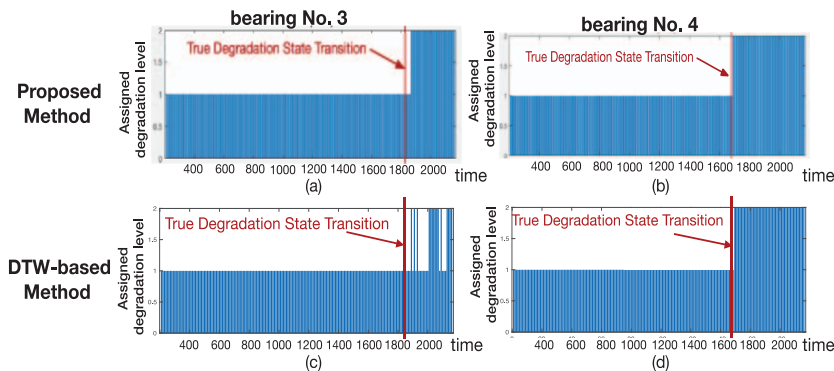


Fig. 4. Clustering of the run-to-failure trajectories of bearings No.3 and 4 obtained by using the proposed (a,b) and *DTW*-based (c,d) methods.

The feature extraction procedure adopted in (Jardine, Lin, and Banjevic 2006), based on the use of six statistical features (mean, root mean square, standard deviation, kurtosis, skewness and min-max range) of the acceleration values measured during the 1-second length window by the sensor closest to the monitored bearing, is used to define the six dimensional input of the proposed method. The Conceptor matrix is computed each time ten 1-second windows are acquired and the corresponding six features are extracted. The first 20 Conceptor matrixes of each run-to-failure trajectory are discarded in order to burn the initialization of the reservoir.

By applying the clustering algorithm in Section 4.1.2, an optimal number of degradation states equal to 2 has been identified. Fig. 3 shows the pairwise Conceptor distance matrix among the time sequences $p_{1:S,1:i}^r$ ($i = 200, 210, \dots, N_r$) extracted from the run-to-

failure trajectory of bearing No.4. It can be seen that there is a clear partition between degradation states 1 (healthy) and 2 (degraded). The vertical and horizontal lines represent the reference transition time. Notice that the separation boundary between degradation states 1 and 2 fits the reference value.

As comparison method, we consider the *DTW*-based spectral clustering with 2 clusters, and non-overlapped and fix-intervalled time windows of length ten 1-second time windows (Cuturi and Blondel 2017). Fig. 4 shows the labels assigned by the proposed and *DTW*-based spectral clustering for bearings No.3 and 4, with the vertical lines indicating the reference values. TABLE 2 reports the comparison of the performance of the proposed method with that of the *DTW*-based spectral clustering applied to the two run-to-failure trajectories, assuming as true transition times the reference times.

The performance of the proposed method is more satisfactory than that of the *DTW*-based spectral clustering. In addition, the *DTW*-based method results are characterized by more missed alarms for bearing No.3 than our proposed method.

TABLE 2. Comparison of the clustering performance of the two training trajectories

Degradation level clustering (Training trajectories)		Accuracy _r	NMI _r
<i>Proposed Method</i>	Bearing No.3	0.97	0.82
	Bearing No.4	0.99	0.91
<i>DTW</i> -based	Bearing No.3	0.90	0.41
	Bearing No.4	0.96	0.80

7. Conclusion

An unsupervised clustering method based on the use of Conceptors has been proposed for data-driven fault diagnostics.

Considering the application to a literature case study containing vibrational data extracted from bearings, the proposed Conceptor-based clustering method has outperformed a traditional clustering approach in terms of accuracy. The obtained results show that Conceptors allow effectively dealing with multivariate time series characterized by long-term temporal dependencies, which are difficult to treat with methods based on the use of sliding time windows. The proposed method contributes to overtaking one of the main limitations toward the practical applications of fault diagnostics, i.e. the need of having available a large amount of labelled data for model training.

Acknowledgement

The work is developed within the research project "SMART MAINTENANCE OF INDUSTRIAL PLANTS AND CIVIL STRUCTURES BY 4.0 MONITORING TECHNOLOGIES AND PROGNOSTIC APPROACHES - MAC4PRO ", sponsored by the call BRIC-2018 of the National Institute for Insurance against Accidents at Work – INAIL. Mingjing Xu gratefully acknowledges the financial support from the China Scholarship Council (No. 201606420061).

References

- Al-Dahidi, S., F. Di Maio, P. Baraldi, E. Zio, and R. Seraoui. 2015. "A Novel Ensemble Clustering for Operational Transients Classification with Application to a Nuclear Power Plant Turbine." *International Journal of Prognostics and Health Management*.
- Al-Dahidi, Sameer, Francesco Di Maio, Piero Baraldi, Enrico Zio, and Redouane Seraoui. 2018. "A Framework for Reconciling Data Clusters from a Fleet of Nuclear Power Plants Turbines for Fault Diagnosis." *Applied Soft Computing Journal*.
<https://doi.org/10.1016/j.asoc.2018.04.044>.
- Baraldi, Piero, Francesco Di Maio, Marco Rigamonti, Enrico Zio, and Redouane Seraoui. 2015. "Clustering for Unsupervised Fault Diagnosis in Nuclear Turbine Shut-down Transients." *Mechanical Systems and Signal Processing* 58: 160–78.
<https://doi.org/10.1016/j.ymssp.2014.12.018>.
- Büsing, Lars, Benjamin Schrauwen, and Robert Legenstein. 2010. "Connectivity, Dynamics, and Memory in Reservoir Computing with Binary and Analog Neurons." *Neural Computation* 22 (5): 1272–1311.
<https://doi.org/10.1162/neco.2009.01-09-947>.
- Cakir, Fatih, Kun He, Sarah Adel Bargal, and Stan Sclaroff. 2019. "Hashing with Mutual Information." *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
<https://doi.org/10.1109/tpami.2019.2914897>.
- Cuturi, Marco, and Mathieu Blondel. 2017. "Soft-DTW: A Differentiable Loss Function for Time-Series." In *34th International Conference on Machine Learning, ICML 2017*.
- Ferreira, Aida A., Teresa B. Ludermir, and Ronaldo R B De Aquino. 2013. "An Approach to Reservoir Computing Design and Training." *Expert Systems with Applications*.
<https://doi.org/10.1016/j.eswa.2013.01.029>.
- Jaeger, Herbert. 2016. "Using Conceptors to Manage Neural Long-Term Memories for Temporal Patterns." *Machine Learning*.
- Jain, Anil K. 2010. "Data Clustering: 50 Years beyond K-Means." *Pattern Recognition Letters* 31 (8): 651–66.
<https://doi.org/10.1016/j.patrec.2009.09.011>.
- Jardine, Andrew K S, Daming Lin, and Dragan Banjevic. 2006. "A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance." *Mechanical Systems and Signal Processing*.
<https://doi.org/10.1016/j.ymssp.2005.09.012>.
- Liu, Han, Jianzhong Zhou, Yanhe Xu, Yang Zheng, Xuanlin Peng, and Wei Jiang. 2018. "Unsupervised Fault Diagnosis of Rolling Bearings Using a Deep Neural Network Based on Generative Adversarial Networks." *Neurocomputing*.
<https://doi.org/10.1016/j.neucom.2018.07.034>.
- Lukoševičius, Mantas, and Herbert Jaeger. 2009. "Reservoir Computing Approaches to Recurrent Neural Network Training." *Computer Science Review* 3 (3): 127–49.
<https://doi.org/10.1016/j.cosrev.2009.03.005>.
- Lukoševičius, Mantas. 2012. "A Practical Guide to Applying Echo State Networks," 659–86.
https://doi.org/10.1007/978-3-642-35289-8_36.

- Luxburg, Ulrike Von. 2007. "A Tutorial on Spectral Clustering." *Statistics and Computing*.
<https://doi.org/10.1007/s11222-007-9033-z>.
- Maio, Francesco Di, Piero Baraldi, Enrico Zio, and Redouane Seraoui. 2013. "Fault Detection in Nuclear Power Plants Components by a Combination of Statistical Methods." *IEEE Transactions on Reliability*.
<https://doi.org/10.1109/TR.2013.2285033>.
- Martins, J. F., V. Fernão Pires, and A. J. Pires. 2007. "Unsupervised Neural-Network-Based Algorithm for an on-Line Diagnosis of Three-Phase Induction Motor Stator Fault." *IEEE Transactions on Industrial Electronics* 54 (1): 259–64.
<https://doi.org/10.1109/TIE.2006.888790>.
- Qian, Guangwu, and Lei Zhang. 2017. "A Simple Feedforward Convolutional Conceptor Neural Network for Classification." *Applied Soft Computing Journal*, 2017.
<https://doi.org/10.1016/j.asoc.2017.08.016>.
- Qiao, Junfei, Fanjun Li, Honggui Han, and Wenjing Li. 2016. "Growing Echo-State Network With Multiple Subreservoirs." *IEEE Transactions on Neural Networks and Learning Systems* 28 (2): 391–404.
<https://doi.org/10.1109/TNNLS.2016.2514275>.
- Qiu, Hai, Jay Lee, Jing Lin, and Gang Yu. 2006. "Wavelet Filter-Based Weak Signature Detection Method and Its Application on Rolling Element Bearing Prognostics." *Journal of Sound and Vibration*.
<https://doi.org/10.1016/j.jsv.2005.03.007>.
- Rousseeuw, Peter J. 1987. "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis." *Journal of Computational and Applied Mathematics*.
[https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- Saravanan, N., and K. I. Ramachandran. 2010. "Incipient Gear Box Fault Diagnosis Using Discrete Wavelet Transform (DWT) for Feature Extraction and Classification Using Artificial Neural Network (ANN)." *Expert Systems with Applications*.
<https://doi.org/10.1016/j.eswa.2009.11.006>.
- Uw, S, Andrew Y Ng, Michael I Jordan, and Yair Weiss. 2001. "On Spectral Clustering: Analysis and an Algorithm." *Advances in Neural Information Processing Systems 14*, 849–56.
<https://doi.org/10.1.1.19.8100>.
- Xu, Mingjing, Piero Baraldi, Sameer Al-Dahidi, and Enrico Zio. 2020a. "Fault Prognostics by an Ensemble of Echo State Networks in Presence of Event Based Measurements." *Engineering Applications of Artificial Intelligence*.
<https://doi.org/10.1016/j.engappai.2019.103346>.
- Xu, Mingjing, Piero Baraldi, Sameer Al-Dahidi, and Enrico Zio. 2020b. "Fault Prognostics in Presence of Event-Based Measurements." In *Proceedings of the 29th European Safety and Reliability Conference, ESREL 2019*. https://doi.org/10.3850/978-981-11-2724-3_0372-cd.
- Ye, Jun, Guojun Qi, Naifan Zhuang, Hao Hu, and Kien A. Hua. 2018. "Learning Compact Features for Human Activity Recognition via Probabilistic First-Take-All." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.
<https://doi.org/10.1109/TPAMI.2018.2874455>.
- Yildiz, Izzet B., Herbert Jaeger, and Stefan J. Kiebel. 2012. "Re-Visiting the Echo State Property." *Neural Networks* 35: 1–9.
<https://doi.org/10.1016/j.neunet.2012.07.005>.
- Yu, Jianbo. 2012. "Health Condition Monitoring of Machines Based on Hidden Markov Model and Contribution Analysis." *IEEE Transactions on Instrumentation and Measurement*.
<https://doi.org/10.1109/TIM.2012.2184015>.
- Zheng, Yajing, Shixin Li, Rui Yan, Huajin Tang, and Kay Chen Tan. 2018. "Sparse Temporal Encoding of Visual Features for Robust Object Recognition by Spiking Neurons." *IEEE Transactions on Neural Networks and Learning Systems*.
<https://doi.org/10.1109/TNNLS.2018.2812811>.