

Fault Detection via modified Principal Direction Divisive Partitioning and application to aerospace electro-mechanical actuators

Mirko Mazzoleni, Simone Formentin, Fabio Previdi and Sergio M. Savaresi

Abstract—In this paper, the use of the Principal Direction Divisive Partitioning (PDDP) method for unsupervised learning is discussed and analyzed with a focus on fault detection applications. Specifically, a geometric limit of the standard algorithm is highlighted by means of a simulation example and a modified version of PDDP is introduced. Such a method is shown to correctly perform data clustering also when the standard algorithm fails. The modified strategy is based on the use of a Chi-squared statistical test and offers more guarantees in terms of detection of a wrong functioning of the system. The proposed algorithm is finally experimentally tested on a fault detection application for aerospace electro-mechanical actuators, for which a comparison with k-means and fuzzy k-means approaches is also provided.

I. INTRODUCTION

Fault Detection and Isolation (FDI) methods aim at monitoring a system, identifying when a fault has occurred, and pinpointing the type and the location of the fault.

FDI approaches can be broadly divided in two classes: model-based and model-free approaches [1]. The former set of methods relies on a mathematical model of the system under consideration, and the detection of the fault is mainly based on the mismatch between the measurements and the state of the system estimated using the model information. Instead, model-free methods do not require the direct estimation of a model of a plant, since the faults are detected mainly based on the comparison of current measurements with processed past data. In this context, pattern-recognition techniques [2] showed to be very useful to derive the required information from data without the need of parameterizing any model.

In machine learning, pattern recognition is concerned with the classification of objects (patterns) into a number of categories or classes. In order to build such classes, after collecting raw data and measurements - which are typically highly redundant - feature extraction is performed to isolate and use only the significant information. This step is very critical because it is representative of an important trade-off between the information which is thrown away, the computational tractability and the ability to generalize to other datasets of the resulting classes [3]. A review of pattern recognition techniques for fault detection is given in [4],

whereas a panoramic view of features and classifiers for fault diagnosis is employed in [5].

The Principal Direction Divisive Partitioning (PDDP) [6] is among the most popular methods for pattern recognition. It was originally proposed to solve a text document classification, and various research activities have worked towards improvements in this direction, see, *e.g.*, [7], [8], [9]. The PDDP approach seems very suited for FDI for several reasons. First of all, its computational cost is roughly linear in the number of non-zeros in the feature matrix and only weakly (logarithmic) scaled with the number of generated clusters. Therefore, despite of its simplicity, it has been proven to produce high quality clusters, especially when the dimensionality of the data is high [8]. This is possible by stopping the singular value decomposition (SVD) at the first singular value/vector and makes PDDP significantly less computationally demanding than other widely known methods like, *e.g.*, Latent Semantic Indexing algorithm (LSI) [10], especially if the data-matrix is sparse and the principal singular vector is calculated by resorting to the Lanczos technique [11], [12].

In this paper, it will be shown that FDI using the standard PDDP approach may yield some problems. More specifically, it will be illustrated by means of a motivating simulation example that PDDP may intrinsically split the data along the first principal component even when two clusters are slender and narrow and their length is greater than the distance between their centroids. This fact constitutes a great limit, as it may significantly jeopardize the quality of clustering. To overcome this problem, as a *first contribution* of the paper, a modified version of PDDP - called mPDDP - is proposed, in which the choice of the cluster to split is based on Chi-squared goodness of the data fitting.

It should also be said that many other variations of the original PDDP algorithm have been proposed to enhance its performance. In [13] the authors developed a non-greedy variant of the algorithm, which tries all the possible choices of partition on a specified number of clusters and principal components, by evaluating the variance within the cluster. The choice of the number of clusters, if not specified a priori, is discussed in [14] with the use of a BIC criterion. The works in [15], [16] focus instead on the choice of which cluster to split. However, as far as the authors are aware, none of the above variations of the PDDP has directly dealt with the problem object of this paper.

As a *second contribution* of this paper, the mPDDP algorithm is experimentally applied on FDI of a real Electro-Mechanical Actuator (EMA) for aerospace applications. **The**

This work was supported by Project HOLMES (Health On-Line Monitoring for Electromechanical actuator Safety).

Mirko Mazzoleni and Fabio Previdi are with Dipartimento di Ingegneria, Università degli Studi di Bergamo, 24044 Dalmine (BG), Italy.

Simone Formentin and Sergio M. Savaresi are with Dipartimento di Eletttronica, Informazione e Bioingegneria, Politecnico di Milano, P.zza Leonardo da Vinci, 32, 20133 Milano, Italy. Email to: simone.formentin@polimi.it

constraints of a lightweight, online and reconfigurable algorithm has dictated the choice of an unsupervised algorithm like the PDDP. Real data will be processed with standard and modified PDDP to further motivate the proposed study on a real-world application and to show the potential of the developed method. The considered major EMA failure modes for fault detection have been first identified in [17].

The remainder of the paper is as follows. In Section II, the standard PDDP method is briefly recalled. The proposed modification of the clustering method is illustrated in Section III, where a simulation example is used to show the limits of the standard approach and visually explain the main idea behind the new algorithm. The modified PDDP is then applied on a real EMA setup for aerospace application, where a significant increase in clustering performance is highlighted. The paper is ended by some concluding remarks.

II. PRINCIPAL DIRECTION DIVISIVE PARTITIONING

The task of unsupervised learning is to reveal the organization of patterns into “sensible” clusters (groups). Similar patterns, in the sense of a defined similarity measure, will be grouped into the same cluster by a clustering algorithm.

Many clustering algorithms have been proposed throughout the years. The oldest ones are based on the Basic Sequential Algorithmic Scheme (BSAS), where each new point is said to belong to a group of points, depending on its distance from the existing clusters. Then, several other categories of methods have been proposed: optimization-based, among which the celebrated k -means algorithm [18]; density-based, *e.g.* the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [19] and hierarchical clustering, like Principal Direction Divisive Partitioning (PDDP) [6].

PDDP belongs to another important class of data-processing techniques, that is SVD-based (Singular Value Decomposition) methods, in which the Latent Semantic Indexing algorithm (LSI) [10], and the LSI-related Linear Least Square Fit (LLSF) algorithm [20] are also included. The considered clustering approach is bisecting divisive clustering: the problem to be solved is the splitting of the data matrix $X = [x_1, x_2, \dots, x_N] \in \mathbb{R}^{m \times N}$ (where m is the dimensionality and N is the number of data) into two sub-matrices (or sub-clusters) $X_L \in \mathbb{R}^{m \times N_L}$ and $X_R \in \mathbb{R}^{m \times N_R}$, $N_L + N_R = N$.

PDDP is mainly based on Principal Components Analysis (PCA), thus involving the eigenvector decomposition of the data covariance matrix, or equivalently a singular value decomposition (SVD) of the data matrix after mean centering. The principal trend in data can be considered in two ways. In PCA, the direction of principal trend is taken as the direction in which the variance (or “spread”) of the data is maximum. The second way to define the principal trend is by means of least squares, in which case the trend is along a line L for which the total sum of squares of orthogonal deviations from L is minimal among all lines in \mathbb{R}^m .

The PDDP algorithm is very popular, mainly for its low computation requirements. As a matter of fact, PDDP has a computational cost roughly linear in the number of non-zeros in the feature matrix and it is characterized by a weak scaling, which is logarithmic with the number of generated clusters. Moreover, it provides a “one-shot” deterministic solution, unlike the initialization-dependent solution given, *e.g.*, by the k -means. For the interested reader, a thorough comparative analysis of bisecting k -means and PDDP is given in [21].

The PDDP algorithm can be formalized as follows.

PDDP clustering algorithm

- 1) Compute the centroid w of X , where w a $m \times 1$ vector
- 2) Compute the auxiliary matrix \tilde{X} as $\tilde{X} = X - w \cdot e$, where e is a N -dimensional row vector of ones, namely $e = [1, 1, \dots, 1]$
- 3) Compute the Singular Value Decompositions (SVD) of \tilde{X} , $\tilde{X} = U\Sigma V^T$, where Σ is a diagonal $m \times N$ matrix, and U and V are orthonormal unitary square matrices having dimension $m \times m$ and $N \times N$, respectively
- 4) Take the first column vector of U , $u = U_1$, and divide $X = [x_1, x_2, \dots, x_N]$, into two sub-clusters X_L and X_R , according to the following rule:

$$\begin{cases} x_i \in X_L & u^T(x_i - w) \leq 0 \\ x_i \in X_R & u^T(x_i - w) > 0 \end{cases} \quad (1)$$

- 5) iterate until the desired number of clusters is reached.

III. MODIFIED PDDP BASED ON STATISTICAL DATA TEST

In this section, a limit of the PDDP algorithm in some practical situations is highlighted. Precisely, it will be shown that the splitting along the first principal component is not always the best choice when two clusters are slender and narrow, and their length is greater than the distance between their centroids. To overcome this problem, a modified version of the PDDP method is proposed. In what follows, first the Chi-squared test will be briefly outlined, then the modified version of PDDP method will be described and illustrated by means of a simulation example.

A. Chi-squared goodness of fit test

The chi-squared goodness of fit test performs a statistical test to assess whether the data is drawn from a Gaussian probability density function (pdf). The situations are then two: either the data was drawn from a normal distribution (assumption H_0) or from another distribution (assumption H_1).

Given an histogram, the question is whether it is consistent with a given pdf. If the histogram has k bins, let b_0, b_1, \dots, b_k be the $k+1$ boundaries. So, x belongs to the i -th bin if $b_{i-1} \leq x \leq b_i$, $i = 1 \dots k$. The counts for the i -th bin is defined as o_i . Since there could be many experiments, o_i is a particular outcome of the random variable O_i . The expected counts e_i are computed from the distribution **with** parameters estimated on the data. To measure the discrepancy between

the observed histogram o_i and the histogram e_i computed under the null hypothesis, it is then natural to use

$$\bar{\chi}^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}. \quad (2)$$

Since the observed bin values o_i are outcomes of the random variables O_i , the value $\bar{\chi}^2$ is itself an outcome of the random variable

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - e_i)^2}{e_i}, \quad (3)$$

which is distributed according to a chi-square probability density function with $d = k - 1$ degrees of freedom, $p_{\chi_d^2}(x)$, because of the constraint $\sum_{i=1}^k O_i = n$.

The p -value is defined as the probability of obtaining a statistic test at least as extreme as the one that was actually observed, assuming that the null hypothesis is true:

$$p = \mathbb{P}[\chi^2 \geq \bar{\chi}^2] = \int_{\bar{\chi}^2}^{\infty} p_{\chi_d^2}(x) dx \quad (4)$$

If $p \leq \alpha$, hypothesis H_0 is rejected at a significance level α and the result is right with probability $1 - \alpha$.

B. mPDDP

To exemplify the problem, consider the case depicted in Fig. 1. The projection on the first principal component, according to the PDDP rule, splits the cluster in two groups: the first cluster is composed by the top half of the blue and the red cluster, while the second cluster is composed by the bottom halves. Projecting on the second component, the two resulting clusters are the blue one and the red one, as an external observer would have suggested.

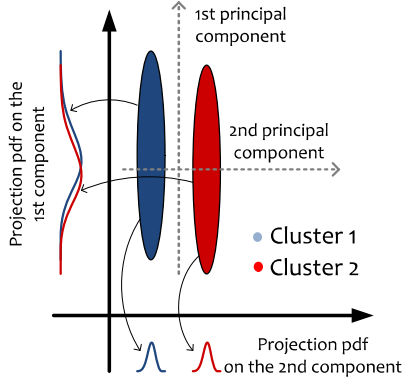


Fig. 1. Dataset projection on the first two principal components

Consider now the simulation example illustrated in Fig. 2. The first and the second big clusters are recognized by the PDDP algorithm and the cluster selected for further splitting is chosen to be the top one. However, at the second step of the algorithm, the red cluster is split into the green and red groups, which is obviously the uncorrect choice.

To address this problem, it should first be noted that the distributions of the data projected on the first component (see again Fig. 1) are significantly overlapped. Moreover, the sum of the two distributions approximates the normal distribution

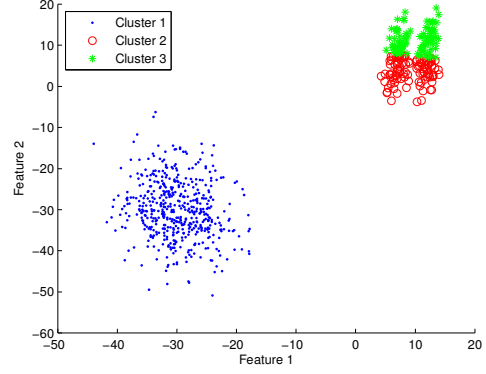


Fig. 2. Clustering produced by the standard algorithm at the second step

better than the sum of the distributions of the projections on the second principal component.

The idea is then to adopt the statistical hypothesis test of the previous subsection to prove if the data really follow a normal distribution. By relying on the previous arguments, the data have to be projected on the direction for which their distribution is less similar to a Gaussian.

Then, a statistical goodness of fit test can be performed on the data projected on each direction. The direction which has generated the data for which the hypothesis test gives the smallest p -value, *i.e.* for which the null hypothesis (the Gaussianity of the data) is rejected, is chosen as the direction where to apply the PDDP. **Notice how the choice of the α level is irrelevant to our analysis, since we rely only on the p -value, regardless of the fact that the test has confirmed or not the null hypothesis. The number L of principal direction to be evaluated is a trade-off between computational complexity and performance. In this application, the choice of L is made once at the beginning, but it can also vary at every step. Investigations about this topic is still ongoing.**

The first three steps of the so-built algorithm are then equal to the old ones, whereas the others need to be reformulated. The overall procedure looks as follows.

mPDDP clustering algorithm

- 1) Compute the centroid w of X , where w is a $m \times 1$ vector
- 2) Compute the auxiliary matrix \tilde{X} as $\tilde{X} = X - w \cdot e$, where e is a N -dimensional row vector of ones, namely $e = [1, 1, \dots, 1]$
- 3) Compute the Singular Value Decompositions (SVD) of \tilde{X} , $\tilde{X} = U \Sigma V^T$, where Σ is a diagonal $m \times N$ matrix, and U and V are orthonormal unitary square matrices having dimension $m \times m$ and $N \times N$, respectively
- 4) Choose the number L of principal components to evaluate, $1 \leq L \leq m$. Take the first L columns of U , $[u_1, u_2, \dots, u_L]$, $l = 1 \dots L$, and compute the data projections onto this principal directions, $z_l = u_l^T (x - w)$.
- 5) Perform a chi-squared goodness of fit test on the z_l vector of data projected, with a significance level α

and compute the p -value p_l , with $l = 1 \dots L$

- 6) Find $j = \underset{l=1 \dots L}{\operatorname{argmin}} p_l$ to find the index of the test which gave the lowest p -value.
- 7) Divide the data into two sub-clusters X_L and X_R , according to the following rule:

$$\begin{cases} x_i \in X_L & u_j^T(x_i - w) \leq 0 \\ x_i \in X_R & u_j^T(x_i - w) > 0 \end{cases} \quad (5)$$

- 8) iterate until the desired number of clusters is reached.

In Fig. 3, the new strategy is applied on the simulation example. Notice that now, as expected, the cluster is split according to the second principal component, unlike using standard PDDP.

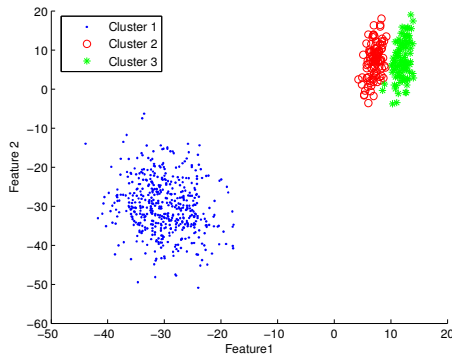


Fig. 3. Clustering produced by the modified algorithm at the second step

IV. APPLICATION TO AEROSPACE ELECTRO-MECHANICAL ACTUATORS

Aerospace applications have very strict constraints on weight and volume, which often penalize redundancy. However, for safety reasons, an actuator with no hardware redundancy must be equipped with a sophisticated diagnostic, prognostic, and recovery system (see [22] for a survey on FDI approaches in aerospace applications). In this application of the technique proposed in this work, the focus will be on Electro-Mechanical Actuators (EMA).

In traditional actuation systems for aerospace, the aerodynamic surfaces are actuated by a series of ballscrews, which are, in turn, actuated by a motor. Each ballscrew moves on a rail, realizing the requested kinematic profile. A **break** or structural failure of the servo motor can lead to an uncontrolled movement of the aerodynamic surfaces, which could be potentially catastrophic. Due to this reason, a couple of actuators (Master and Slave), which are synchronous but independent in term of motion generation and stop, are used to prevent such situations.

The system under test in this work, illustrated in Fig. 4, is composed by two five phases brushless motor (the Master and Slave EMAs) with a ballscrew transmission, and by an hydraulic cylinder aimed to generate a resistive load. The faults to be detected are:

- **One** raceway clogged

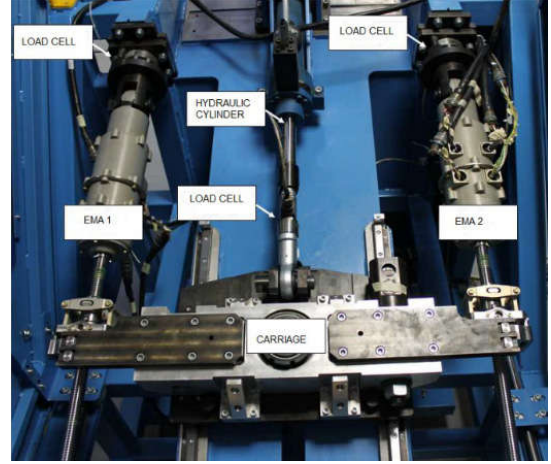


Fig. 4. The experimental setup used in this work.

- **Two** raceways clogged
- worn ball bearings

The training data were collected for different types of input position profiles, with the aim to excite the system in all its components. The type of input profiles was chosen to be a square wave-like signal, with different magnitude and frequency of repetition. For each input, the data were collected for every fault conditions, with a sampling frequency of $f_s = 10kHz$. The variables chosen for the analysis were referred at the Master actuator, and they are:

- 1) Position set-point
- 2) Position measured
- 3) One phase current (phase A)
- 4) Torque-generating current

so, for each input profile, there are 4 variables and 3 fault conditions, for a total of 12 variables/profile. For simplicity, the load profiles were set to zero for each experiment.

Recent studies [23], [24] report that existing current and position/speed sensors equipping aerospace EMA are a promising tool for health monitoring of electromechanical actuators based on screw systems. Therefore, features to be used for FDI are computed based on the current signals from the various fault types and profiles. To take into account the needs for a real-time evaluation, and to make the statistical indexes more reliable, each feature sample is computed via an overlapping moving window, with length of 3s and slide length of 1.5s.

For feature extraction, the motor state transition from one fault to another is sometimes simulated as an abrupt change and some others as a smooth transition.

In this application, up to 21 features have been computed, spanning from time domain, frequency domain, and time-frequency domain. Time domain features include general purpose indexes, like Root Mean Square (RMS) value, skewness, kurtosis, sixth central moment, shape and crest factors, peak-to-valley value, energy operator [25], [26], [5], and application specific indexes, like position error [17] and the torque-speed ratio.

Frequency domain features consist mainly of the magnitude of the Fast Fourier Transform (FFT) over 3 sets of frequencies. Three indexes of these type have been extracted, based on the value of the magnitude at different frequency bands. Other features are mean frequency, frequency center, RMS and standard deviation in the frequency domain [5].

After the computation of the features, any point in the feature space is t -dimensional, with $t = 21$. In order to reduce the computational time and to prevent possible causes of bad generalization, feature selection is performed. Specifically, the method of the Linear Discriminant Analysis (LDA), which goes back to the pioneering work of Fisher [27], is used. The LDA is a supervised method, which means that the class which data belong is given to the algorithm; the dimension reduction is achieved via a linear combinations of the existing features, by seeking the direction w in the t -dimensional space, along which the classes are best separated in some way. This can be done by maximizing the Fisher Discriminant Ratio, which, for the two-classes case, it is equal to:

$$F(w) = \frac{(\mu_1(w) - \mu_2(w))^2}{\sigma_1(w)^2 + \sigma_2(w)^2},$$

being μ_1 and μ_2 the mean of the class one and two, respectively, and with σ_1^2, σ_2^2 the variances of the class one and two. **These parameters are scalar values after the projection along the direction w .** Notice that $F(w)$ is large if the classes are well separated. The method can be straightforwardly extended to be used in the multi-class case, see [28].

The peculiarity of this technique is that it produces a number of feature which is at most equals to the number of classes minus one. So, in this case, since there are 3 classes, after this step the feature space changes from 21-dimensional to 2-dimensional, and the new features, which we can call “Feature1” and “Feature2”, are linear combination of the previous ones.

After the feature extraction and selection phases have been performed, classification is needed. **A point belongs to the cluster at minimum Euclidean distance, computed respectively to the cluster center.** Here, it is possible to compare the standard and modified PDDP algorithms. Fig. 5 and Fig. 6 show the real bounds (solid) and the boundaries found by the clustering algorithms (dashed). The three considered faults are highlighted using three different colors and the misclassified points are put in evidence with surrounding circles. It can be noted that the boundaries found by the new algorithm are closer to the true ones than the stripes selected by the standard algorithm. This fact produces a much better detection rate and a much smaller number of misclassified points, as summarized in Table I. In a safety critical application like this one, this result largely increases the reliability of the overall system.

For the sake of completeness, also the performance of the k-means [29] and the fuzzy k-means [30] algorithms are evaluated. The results with such methods are shown in Fig. 7 and 8, whereas the main quality indices are summarized

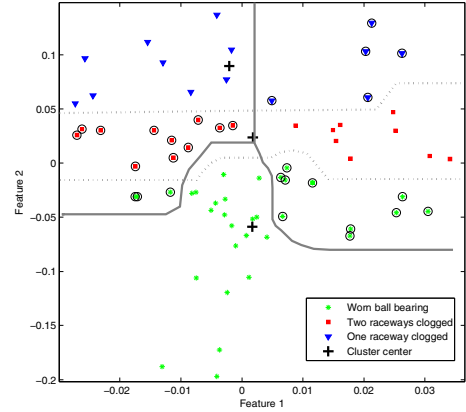


Fig. 5. Performance of the standard PDDP algorithm, true boundaries (solid lines), obtained boundaries (dotted lines), misclassified points (circled)

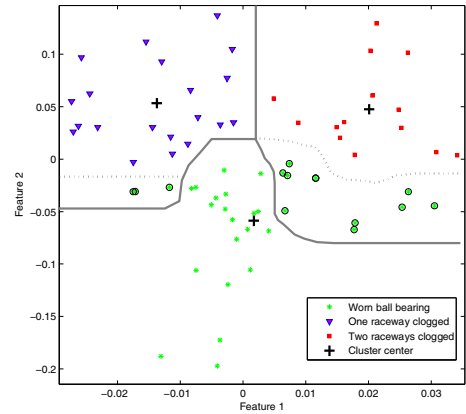


Fig. 6. Performance of the proposed mPDDP algorithm, true boundaries (solid lines), obtained boundaries (dotted lines), misclassified points (circled).

in Table I. Notice that the proposed mPDDP algorithm outperforms also this different clustering approach.

V. CONCLUSIONS

Model-free fault diagnosis and isolation (FDI) is an appealing tool to make a safety critical system reliable against model uncertainties, since no models of the system are used to estimate whether a fault has occurred or not. In particular, model-free FDI using pattern recognition techniques shares the advantages of feature selection and unsupervised learning methods, that is, large datasets can be handled by smart dimensionality reduction and no previous labeling (of the faults) is needed. PDDP is among the most popular techniques in this framework. However, in this work, it is shown that PDDP may provide wrong results in case of particular distribution of the data in the features space. Therefore, a modified version of the PDDP algorithm is proposed, called mPDDP, based on a Chi-squared statistical test. The proposed method showed to be very effective when applied

TABLE I
RESULTS OF CLASSIFICATION

Algorithm	Misclassified	Detection rate
mPDDP	13/67	0.1940
PDDP	29/67	0.4328
k-means	30/67	0.4478
fuzzy k-means	27/67	0.4030

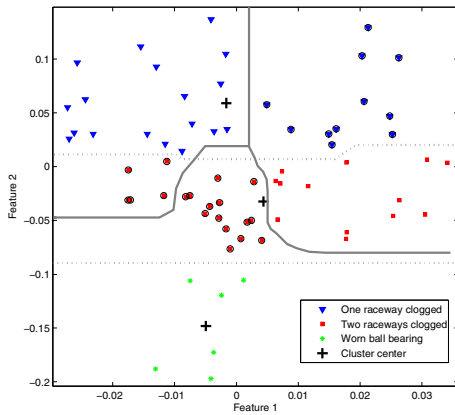


Fig. 7. Performance of the k-means algorithm, true boundaries (solid lines), obtained boundaries (dotted lines), misclassified points (circled).

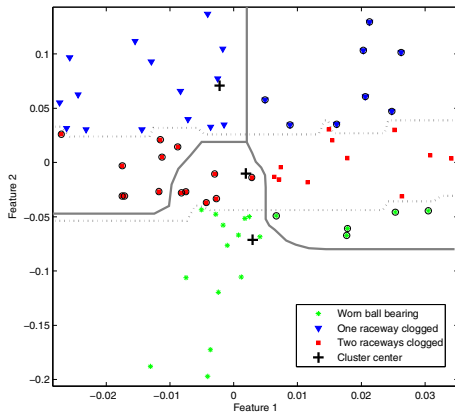


Fig. 8. Performance of the fuzzy k-means algorithm, true boundaries (solid lines), obtained boundaries (dotted lines), misclassified points (circled).

on experimental data taken from an aerospace EMA setup (where suitable faults were simulated ad-hoc).

Future work will be dedicated to the theoretical analysis of the proposed approach, and its integration within a standardized FDI procedure. Moreover, different applications will be addressed to better understand the potential of the proposed approach with respect to other existing methods.

REFERENCES

- [1] Y. Zhang and J. Jiang, "Bibliographical review on reconfigurable fault-tolerant control systems," *Annual reviews in control*, vol. 32, no. 2, pp. 229–252, 2008.
- [2] G. Guglielmi, T. Parisini, and G. Rossi, "Keynote paper: Fault diagnosis and neural networks: A power plant application," *Control Engineering Practice*, vol. 3, no. 5, pp. 601–620, 1995.
- [3] V. Vapnik, *The nature of statistical learning theory*. Springer, 2000.
- [4] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin, "A review of process fault detection and diagnosis: Part iii: Process history based methods," *Computers & chemical engineering*, vol. 27, no. 3, pp. 327–346, 2003.
- [5] Y. Lei, M. J. Zuo, Z. He, and Y. Zi, "A multidimensional hybrid intelligent method for gear fault diagnosis," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1419–1430, 2010.
- [6] D. Boley, "Principal direction divisive partitioning," *Data mining and knowledge discovery*, vol. 2, no. 4, pp. 325–344, 1998.
- [7] D. Boley, M. Gini, R. Gross, E.-H. S. Han, K. Hastings, G. Karypis, V. Kumar, B. Mobasher, and J. Moore, "Partitioning-based clustering for web document categorization," *Decision Support Systems*, vol. 27, no. 3, pp. 329–341, 1999.
- [8] D. Boley, "Hierarchical taxonomies using divisive partitioning," tech. rep., Technical Report TR-98-012, Department of Computer Science, University of Minnesota, Minneapolis, 1998.
- [9] E.-H. Han, D. Boley, M. Gini, R. Gross, K. Hastings, G. Karypis, V. Kumar, B. Mobasher, and J. Moore, "Webcase: a web agent for document categorization and exploration," in *Proceedings of the second international conference on Autonomous agents*, pp. 408–415, ACM, 1998.
- [10] S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman, "Indexing by latent semantic analysis," *JASIS*, vol. 41, no. 6, pp. 391–407, 1990.
- [11] C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*. United States Government Press Office, 1950.
- [12] G. H. Golub and C. F. Van Loan, *Matrix computations*, vol. 3. JHU Press, 2012.
- [13] M. Nilsson, "Hierarchical clustering using non-greedy principal direction divisive partitioning," *Information Retrieval*, vol. 5, no. 4, pp. 311–321, 2002.
- [14] C. Kruengkrai, V. Sornlertlamvanich, and H. Isahara, "Refining a divisive partitioning algorithm for unsupervised clustering," in *HIS*, pp. 535–542, 2003.
- [15] S. M. Savaresi, D. L. Boley, S. Bittanti, and G. Gazzaniga, "Choosing the cluster to split in bisecting divisive clustering algorithms," tech. rep., Technical Report TR-00-53, 2002.
- [16] S. Tasoulis and D. Tasoulis, "Improving principal direction divisive clustering," in *14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008), Workshop on Data Mining using Matrices and Tensors, Las Vegas, USA, 2008*.
- [17] E. Balaban, P. Bansal, P. Stoelting, A. Saxena, K. F. Goebel, and S. Curran, "A diagnostic approach for electro-mechanical actuators in aerospace systems," in *Aerospace conference, 2009 IEEE*, pp. 1–13, IEEE, 2009.
- [18] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, p. 14, California, USA, 1967.
- [19] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *KDD*, vol. 96, pp. 226–231, 1996.
- [20] C. G. Chute and Y. Yang, "An overview of statistical methods for the classification and retrieval of patient events," *Methods of information in medicine*, vol. 34, no. 1–2, pp. 104–110, 1995.
- [21] S. M. Savaresi and D. L. Boley, "A comparative analysis on the bisecting k-means and the pddp clustering algorithms," *Intelligent Data Analysis*, vol. 8, no. 4, pp. 345–362, 2004.
- [22] J. Marzat, H. Piet-Lahanier, F. Damongeot, and E. Walter, "Model-based fault diagnosis for aerospace systems: a survey," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, vol. 226, no. 10, pp. 1329–1360, 2012.
- [23] A. Isturiz, J. Vinals, J. M. Abete, and A. Iturrospe, "Health monitoring strategy for electromechanical actuator systems and components. screw backlash and fatigue estimation," in *Recent Advances in Aerospace Actuation Systems and Components*, vol. 5, 2012.
- [24] A. Picot, Z. Obeid, J. Régner, S. Poignant, O. Darnis, and P. Mausson, "Statistic-based spectral indicator for bearing fault detection in permanent-magnet synchronous machines using the stator current," *Mechanical Systems and Signal Processing*, 2014.
- [25] T. W. Rauber, E. M. do Nascimento, E. D. Wandekokem, and F. M. Varejão, "Pattern recognition based fault diagnosis in industrial processes: Review and application," in *Pattern Recognition Recent Advances* (A. Herout, ed.), ch. 25, Springer, 2010.
- [26] J. Zarei, "Induction motors bearing fault detection using pattern recognition techniques," *Expert systems with Applications*, vol. 39, no. 1, pp. 68–73, 2012.
- [27] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.
- [28] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Academic Press, 2009.
- [29] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.
- [30] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," *IEEE Transactions on Systems, Man and Cybernetics*, no. 4, pp. 580–585, 1985.