

An approach to robust fault diagnosis in mechanical systems using computational intelligence

Adrián Rodríguez Ramos¹ · José M. Bernal de Lázaro¹ · Alberto Prieto-Moreno¹ · Antônio José da Silva Neto² · Orestes Llanes-Santiago¹ 

Received: 13 February 2017 / Accepted: 2 July 2017 / Published online: 10 July 2017
© Springer Science+Business Media, LLC 2017

Abstract In this paper a novel approach to design robust fault diagnosis systems in mechanical systems using historical data and computational intelligence techniques is presented. First, the pre-processing of the data to remove the outliers is performed with the aim of reducing the classification errors. To accomplish this objective, the Density Oriented Fuzzy C-Means (DOFCM) algorithm is used. Later on, the Kernel Fuzzy C-Means (KFCM) algorithm is used to achieve greater separability among the classes, and reducing the classification errors. Finally, an optimization process of the parameters used in the training state by the DOFCM and KFCM for improving the classification results is developed using the bioinspired algorithm Ant Colony Optimization. The proposal was validated using the DAMADICS (Development and Application of Methods for Actuator Diagnosis in Industrial Control Systems) benchmark. The satisfactory results obtained indicate the feasibility of the proposal.

Keywords Robust fault diagnosis · Mechanical systems · Computational intelligence · Fuzzy clustering techniques · Optimal parameters

Introduction

In modern industries there are higher and increasing requirements associated with the efficiency of processes, the quality of the products and the fulfillment of environmental and industrial safety regulations (Hwang et al. 2010; Venkatasubramanian et al. 2003a).

Mechanical systems exist in almost all manufacturing industries, and a large part of the faults that occur in these industries are associated with this type of systems (Aydin et al. 2014). In general, the faults have an unfavorable impact in the productivity, the environment and the safety of operators. In an industrial context, safety is associated with a set of specifications or standards that manufacturers must meet in order to reduce the accident risks. With this purpose, it is important to incorporate automatic control and supervisory systems into industrial processes, allowing satisfactory operation of these through compensating the effects of perturbations and changes that might occur in them. Therefore, in order to guarantee that the operation of a system satisfies performance specifications, the faults need to be detected and isolated, being these tasks associated to the fault diagnosis systems (Isermann 2011).

In general form, the fault diagnosis methods can be classified into two categories: models-based methods (Camps Echeverría et al. 2014b, a; Ding 2008; Patan 2008; Venkatasubramanian et al. 2003a, b) and process history-based methods (Fan and Wang 2014; Bernal de Lázaro et al. 2016, 2015; Pang et al. 2014; Sina et al. 2014). In the first approach, the diagnosis tools use models which describe the operation

✉ Orestes Llanes-Santiago
orestes@tesla.cujae.edu.cu

Adrián Rodríguez Ramos
adrian.rr@automatica.cujae.edu.cu

José M. Bernal de Lázaro
jbernal@automatica.cujae.edu.cu

Alberto Prieto-Moreno
albprieto@automatica.cujae.edu.cu

Antônio José da Silva Neto
ajsneto@iprj.uerj.br

¹ Departamento de Automática y Computación, Universidad Tecnológica de la Habana “José Antonio Echeverría” (CUJAE), calle 114 No. 11901, CUJAE, Marianao, CP 19390 Havana, Cuba

² Instituto Politécnico da Universidade do Estado do Rio de Janeiro (IPRJ/UERJ), Rua Bonfim, 25 - Campus UERJ, Nova Friburgo, RJ 28625-570, Brazil

of the processes. These tools are based on the residue generation obtained from the difference between the measured variables in the real process, and the values of the same variables obtained from the model. This type of methods entails an elevated knowledge about the characteristics of the processes, their parameters, and operation zones. However, it is usually very difficult to achieve due to the current complexity of the industrial processes. In mechanical systems there are several applications where these techniques have been used (Karami et al. 2010; Kourad et al. 2012).

On the other hand, the approaches based in historical data do not need a mathematical model, and they do not require much prior knowledge of the process parameters (Choudhary et al. 2008; Wang and Hu 2009). These characteristics constitute an advantage for complex systems, where relationships among variables are nonlinear, not totally known, and therefore, it is very difficult to obtain an analytical model that describes efficiently the dynamics of the process. In the case of mechanical systems, some techniques has been used to fault diagnosis. For example, Motor Current Signature Analysis (MCSA) is the most widely used method to detect various motor faults (Sharifi and Ebrahimi 2011). In order to extract fault features of large-scale power equipment from strong background noise, a fault diagnosis method based on the Wavelet de-noising was proposed (Liu et al. 2016), and broken rotor bar faults were detected using a nonlinear time series analysis (Silva et al. 2008).

Among the various techniques used in fault diagnosis of mechanical systems the use of computational intelligence tools as Neural networks (Hou et al. 2003), Support Vector Machines (Hu et al. 2007), and Fuzzy logic (Bocaniala et al. 2005; Rodríguez Ramos et al. 2016) are emphasized. In addition, there has been a significant increase in the use of the fuzzy clustering methods in recent years (Bedoya et al. 2012; Botia et al. 2013; Jahromi et al. 2016; Seera et al. 2015; Xu et al. 2016).

Fuzzy clustering techniques are very important tools of unsupervised data classification (Gosain and Dahika 2016). They can be used to organize data into groups based on similarities among the individual data. Fuzzy clustering deals with the uncertainty and vagueness existing in a wide variety of applications, as for example: image processing, pattern recognition, object recognition, modeling and identification (Jiang et al. 2016; Kesemen et al. 2016; Leski 2016; Saltos and Weber 2016; Thong and Son 2016b; Vonga et al. 2014; Zhang et al. 2016). The main focus of all fuzzy clustering techniques is to improve the clustering by avoiding the influence of the noise and outlier data.

The Fuzzy C-Means (FCM) algorithm (Bezdek 1981), is one of the most widely used algorithm for clustering due to its satisfactory results for overlapped data. Unlike k-means algorithm, it considers the possible membership of the data points to more than one cluster. FCM algorithm obtains very

good results with noise free data but are highly sensitive to noisy data and outliers (Gosain and Dahika 2016).

Other similar techniques are Possibilistic C-Means (PCM) (Krishnapuram and Keller 1993) and Possibilistic Fuzzy C-Means (PFCM) (Pal et al. 2005). They analyze each cluster as a possibilistic partition. However, PCM fails to find optimal clusters in the presence of noise (Gosain and Dahika 2016) and PFCM does not yield satisfactory results when data set consists of two clusters which are highly unlike in size and outliers exist (Gosain and Dahika 2016; Kaur et al. 2013). Noise Clustering (NC) (Dave 1991; Dave and Krishnapuram 1997), Credibility Fuzzy C-Means (CFCM) (Chintalapudi and Kam 1998), and Density Oriented Fuzzy C-Means (DOFCM) (Kaur 2011) algorithms were proposed specifically to work efficiently with noisy data.

The clustering output depends upon various parameters such as distribution of data points inside and outside the cluster, shape of the cluster and linear or non-linear separability. The effectiveness of the clustering method strongly relies on the choice of the metric distance adopted. FCM uses Euclidean distance as the distance measure, and therefore, it can only be able to detect hyper spherical clusters. Researchers have proposed other distance measures such as, Mahalanobis distance measure, and Kernel based distance measure in data space and in high dimensional feature space, such that non-hyper spherical/non-linear clusters can be detected (Zhang and Chen 2003, 2004).

Another common problem of fuzzy clustering methods is that their performance depend significantly on the initialization of their parameters. In many occasions, it is necessary to make multiple runs of the algorithm in order to obtain good results which is time consuming, and not always the obtaining of the best solution is guaranteed.

In order to overcome these problems, in this paper a new fault diagnosis methodology in mechanical systems using fuzzy clustering techniques is proposed. The methodology consists of three basic steps. First, the pre-processing of data to remove outliers is performed. To achieve this goal the DOFCM algorithm is used. Second, the classification process is developed. For this, the Kernel Fuzzy C-means (KFCM) algorithm is used to obtain a better separability among classes and therefore, the classification results are improved. Finally, a third step is used to optimize the parameters m (factor that regulates the fuzziness of the resulting partition) and σ (bandwidth and indicates the degree of smoothness of the Gaussian kernel function) of the algorithms used in the previous stages using Ant Colony Optimization (ACO) algorithm.

The main contribution of this paper is the obtaining of a robust fault diagnosis scheme in mechanical systems, that adequately combines fuzzy clustering algorithms to solve the drawbacks of this type of technique when the data is affected by noise and outliers, and improving the classification by

using kernel tools whose parameters are optimized to obtain the best results.

The organization of the paper is as follows: in “General description of the principal tools used in the proposal” section is presented the general characteristics of the tools used in the proposed methodology. In “Proposal of classification methodology using computational intelligence tools” section, a description of the new classification methodology using fuzzy clustering techniques is presented. The “Benchmark case study: DAMADICS” section presents the case study used to validate the proposed methodology, as well as the experiment design. In “Analysis and discussion of results” section, an analysis of the obtained results is performed. A comparison with recent fuzzy clustering algorithms is performed in “Comparison with other fuzzy clustering algorithms” section. Finally, the conclusions are presented.

General description of the principal tools used in the proposal

Density Oriented Fuzzy C-Means (DOFCM)

The algorithm attempts to decrease the noise sensitivity in fuzzy clustering by identifying outliers before the clustering process. The DOFCM algorithm creates $c + 1$ clusters with c good clusters and one cluster of noise. This algorithm identifies outliers before the construction of the clusters, based on the density of data set, as it is shown in Fig. 1.

The neighborhood of a given radius of each point in a data set has to contain at least a minimum number of other points. DOFCM defines a density factor, called the neighborhood membership, which express the measure density of an object in relation to its neighborhood. The neighborhood membership of a point i in X is defined as:

$$M_{neighborhood}^i = \frac{\eta_{neighborhood}^i}{\eta_{max}} \tag{1}$$

where $\eta_{neighborhood}^i$ is the number of points in the point neighborhood i ; η_{max} is the maximum number of points in the neighborhood of any point in the data set.

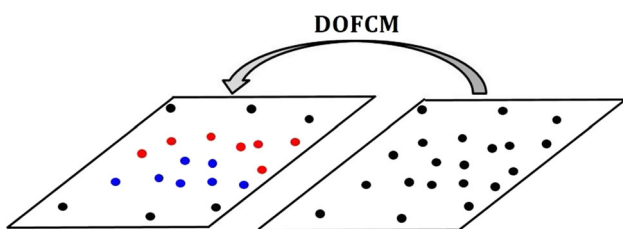


Fig. 1 Identification of outliers with the DOFCM algorithm

If the point q is in the point neighborhood of the point i , q will satisfy:

$$q \in X | dist(i, q) \leq r_{neighborhood} \tag{2}$$

where $r_{neighborhood}$ is the radius of neighborhood, and $dist(i, q)$ is the distance between points i and q . Calculation of neighborhood radius is done in the similar way to (Ester et al. 1996).

Neighborhood membership of each point in the data set X is calculated using Eq. (1). The threshold value α is selected from the complete range of neighborhood membership values, depending on the density of points in the data set. The point will be considered as an outlier if its neighborhood membership is less than α . Let i be a point in the data set X , then

$$\begin{cases} M_{neighborhood}^i < \alpha \text{ then } i \text{ outlier} \\ M_{neighborhood}^i \geq \alpha \text{ then } i \text{ non-outlier} \end{cases} \tag{3}$$

α can be selected from the range of $M_{neighborhood}^i$ values after observing the density of points in the data set and it should be close to zero. Ideally, a point will be classified as outlier only if there is no other point in its neighborhood, i.e., when neighborhood membership is zero or threshold value $\alpha = 0$. However, in this scheme, a point is considered as an outlier when its neighborhood membership is less than α , where α is a critical parameter to identify the outlier points. Its value depends upon the nature of data set, i.e., taking into account the density of the data set, then, its value will vary for different data sets.

After identifying the outliers, the process of clustering begins. DOFCM reformulates FCM objective function as:

$$J_{DOFCM}(X; U, v) = \sum_{i=1}^{c+1} \sum_{k=1}^N (\mu_{ik})^m (d_{ik})^2 \tag{4}$$

where, the distances are defined by,

$$d_{ik}^2 = (\mathbf{x}_k - \mathbf{v}_i)^T \mathbf{A}_i (\mathbf{x}_k - \mathbf{v}_i), \forall k, i = 1 \dots c \tag{5}$$

Membership function μ_{ik} is modified as:

$$\mu_{ik} = \begin{cases} \frac{1}{\sum_{j=1}^c (d_{ik}/d_{jk})^{2/(m-1)}} & \text{if non-outlier} \\ 0 & \text{if outlier} \end{cases} \tag{6}$$

To update the centroid, DOFCM algorithm uses Eq. (7) as FCM algorithm. For the constraint on fuzzy membership, DOFCM algorithm uses Eq. (8). The DOFCM algorithm is presented in Algorithm 1.

$$\mathbf{v}_i = \frac{\sum_{k=1}^N (\mu_{ik}^m \mathbf{x}_k)}{\sum_{k=1}^N \mu_{ik}^m} \tag{7}$$

$$0 \leq \sum_{i=1}^c \mu_{ik} \leq 1, k = 1, 2, \dots, N \tag{8}$$

Algorithm 1 Density Oriented Fuzzy C-Means (DOFCM)

Input: data with outliers \mathbf{X} , $c, \epsilon > 0, m > 1, Itr_max$
Output: data without outliers \mathbf{X}_p
Identification of the outliers:
 Calculate neighborhood radius.
 Calculate $\eta_{neighborhood}^i$ according to (2).
 Select η_{max} .
 Calculate $M_{neighborhood}^i$ according to (1).
 With the given value of α , identify outliers according to (3).
Clustering process:
 Initialize \mathbf{U} to random fuzzy partition.
for $l = 1$ to $l = Itr_max$ **do**
 Update classes centers according to (7).
 Calculate the distance d_{ik} according to (5).
 Update \mathbf{U} according to (6).
 Verify stopping criterion.
end for

Kernel Fuzzy C-Means (KFCM)

KFCM represents the kernel version of FCM. This algorithm uses a kernel function for mapping the data points from the input space to a high dimensional space, as it is shown in Fig. 2.

KFCM algorithm modifies the objective function of FCM using the mapping Φ as:

$$J_{KFCM} = \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m \|\Phi(\mathbf{x}_k) - \Phi(\mathbf{v}_i)\|^2 \tag{9}$$

subject to:

$$\sum_{i=1}^c \mu_{ik} = 1, k = 1, 2, \dots, N \tag{10}$$

where $\|\Phi(\mathbf{x}_k) - \Phi(\mathbf{v}_i)\|^2$ is the square of the distance between $\Phi(\mathbf{x}_k)$ and $\Phi(\mathbf{v}_i)$. The distance in the feature space is calculated through the kernel in the input space as follows:

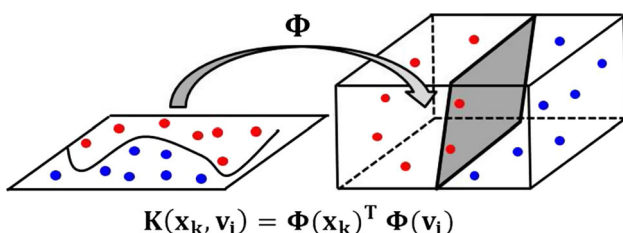


Fig. 2 KFCM feature space and kernel space

$$\|\Phi(\mathbf{x}_k) - \Phi(\mathbf{v}_i)\|^2 = \mathbf{K}(\mathbf{x}_k, \mathbf{x}_k) - 2\mathbf{K}(\mathbf{x}_k, \mathbf{v}_i) + \mathbf{K}(\mathbf{v}_i, \mathbf{v}_i) \tag{11}$$

If the Gaussian kernel is used, then $\mathbf{K}(\mathbf{x}, \mathbf{x}) = \mathbf{1}$ and $\|\Phi(\mathbf{x}_k) - \Phi(\mathbf{v}_i)\|^2 = 2(1 - \mathbf{K}(\mathbf{x}_k, \mathbf{v}_i))$. Thus Eq. (4) can be written as:

$$J_{KFCM} = 2 \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m \|1 - \mathbf{K}(\mathbf{x}_k, \mathbf{v}_i)\|^2 \tag{12}$$

where,

$$\mathbf{K}(\mathbf{x}_k, \mathbf{v}_i) = e^{-\|\mathbf{x}_k - \mathbf{v}_i\|^2 / \sigma^2} \tag{13}$$

Minimizing Eq. (12) under the constraint shown in Eq. (10), yields:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{1 - \mathbf{K}(\mathbf{x}_k, \mathbf{v}_j)}{1 - \mathbf{K}(\mathbf{x}_k, \mathbf{v}_i)} \right)^{1/(m-1)}} \tag{14}$$

$$\mathbf{v}_i = \frac{\sum_{k=1}^N (\mu_{ik}^m \mathbf{K}(\mathbf{x}_k, \mathbf{v}_i) \mathbf{x}_k)}{\sum_{k=1}^N \mu_{ik}^m \mathbf{K}(\mathbf{x}_k, \mathbf{v}_i)} \tag{15}$$

The KFCM algorithm is presented in Algorithm 2.

Algorithm 2 Kernel Fuzzy C-Means (KFCM)

Input: data without outliers \mathbf{X}_p , $c, \epsilon > 0, m > 1, \sigma, Itr_max$.
Output: fuzzy partition \mathbf{U} , class centers \mathbf{V} .
 Initialize \mathbf{U} to random fuzzy partition.
for $l = 1$ to $l = Itr_max$ **do**
 Update the centers of classes according to (15) for Gaussian kernels.
 Calculate the distances according to (11).
 Update \mathbf{U} according to (14).
 Verify stopping criterion.
end for

Proposal of classification methodology using computational intelligence tools

The classification scheme proposed in this paper is shown in Fig. 3. It presents an off-line learning or training stage and an online recognition stage. In the training stage the historical data of the process are used to train (modeling the functional stages through the clusters) the fuzzy classifier. After the training, the classifier is used online (recognition) in order to process every new sample taken from the process. The result intends to offer information about the system state to the operator in real-time.

The clustering methods create the classes based on a measure of similitude by bringing together the data acquired

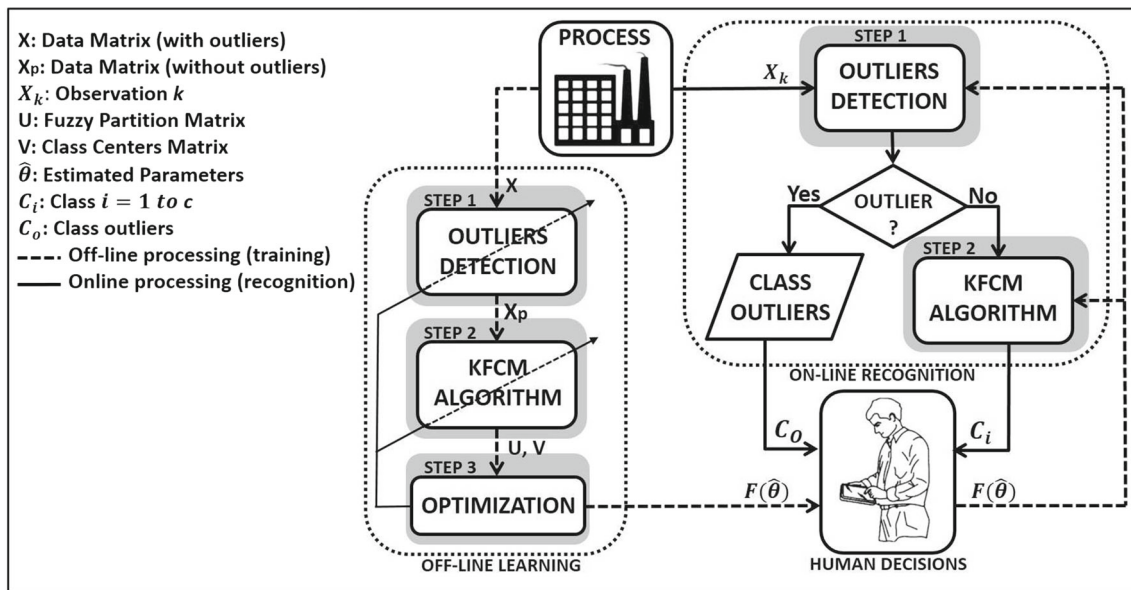


Fig. 3 Classification scheme using fuzzy clustering

by a Supervisory Control and Data Acquisition System (SCADA). These classes can be associated to functional states. When fuzzy classifiers are used in the classification process, each sample is compared with the center of each class using a measure of similitude to determine the membership degree of the sample to each class. In general, the highest membership degree determines the class to which the sample is assigned, as it is showed in Eq. (16).

$$C_i = \{j : \max \{\mu_{ijk}\}, \forall i, k\} \tag{16}$$

Off-line training

In the first step, the center of each known classes $v = v_1, v_2, \dots, v_N$ is determined by using a historical data set representative of the different operation states of the process. A set of *N* observations (data points) $X = [x_1, x_2, \dots, x_N]$ are classified into *c*+1 groups or classes using the DOFCM algorithm. The *c* classes represent the normal operation conditions (NOC) of the process, and the faults to be diagnosed. They contain the information to be used in the next step. The other remaining class contains the data points identified as outliers by the DOFCM algorithm, and they are not used in the next step.

In the second step, the KFCM algorithm receives the set of observations classified by the DOFCM algorithm in the *c* classes. The KFCM algorithm maps these observations into a higher dimensional space in which the classification process obtains better results of satisfactory classifications. The Fig. 4 shows the procedure described in steps 1 and 2.

Finally, a third step to optimize the parameters of the algorithms used in steps 1 and 2 is implemented. In this step, the

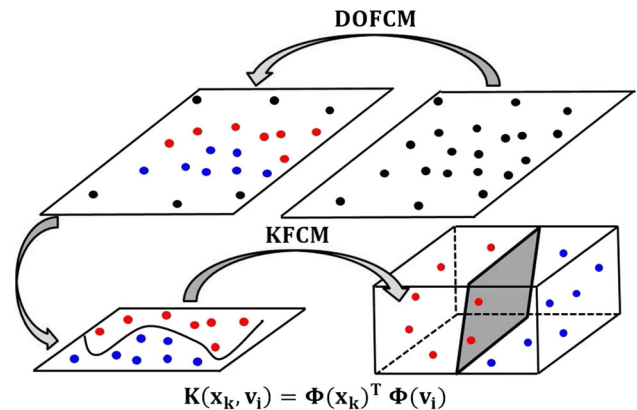


Fig. 4 Procedure performed by the DOFCM and KFCM algorithms

parameters *m* and σ are estimated to optimize a validity index using an optimization algorithm. This will allow to obtain an improved *U* partition matrix, and therefore, a better position of the centers of the classes that characterize the different operation states of the system. Later, the estimated values of *m* in Eqs. (4, 12) and σ in Eq. (13) will be used during the online recognition, and it will contribute to improve the classification of the samples obtained by the data acquisition system from the system.

The validity measures are indexes allowing to evaluate quantitatively the result of a clustering method and comparing its behavior when its parameters vary. Some indexes evaluate the resulting *U* matrix, while others are focused on the geometric resulting structure. The partition coefficient (*PC*) (Li et al. 2012; Pakhira et al. 2004; Wu and Yang 2005), which measures the fuzziness degree of the partition *U*, is

used as validity measure in this case. Its expression is shown in the Eq. (17).

$$PC = \frac{1}{N} \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^2 \quad (17)$$

If the partition U is less fuzzy, the clustering process is better. Being analyzed in a different way, it allows to measure the degree of overlapping among the classes. In this case, the optimum comes up when PC is maximized, i.e., when each pattern belongs to only one group. Likewise, minimum comes up when each pattern belongs to each group.

Therefore, the optimization problem is defined as:

$$\max \{PC\} = \frac{1}{N} \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^2$$

subject to:

$$m_{min} < m \leq m_{max}$$

$$\sigma_{min} \leq \sigma \leq \sigma_{max}$$

In many scientific areas, and in particular in the fault diagnosis field, bio-inspired algorithms have been widely used with excellent results (Camps Echevarría et al. (2010); Liu and Lv 2009; Lobato et al. 2009) to solve optimization problems. They can efficiently locate the neighborhood of the global optimum in most occasions with an acceptable computational time. There is a large number of bio-inspired algorithms, in their original and improved versions. Some examples are Genetic Algorithm (GA), Differential Evolution (DE), Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) among others. In this proposal, the typical ACO algorithm was used to obtain the optimum values of the parameters m and σ after a comparison with PSO and DE algorithms.

On-line recognition

In this stage the fuzzy clustering algorithms are modified and the updating of the center of each class is not developed. The principal reason for doing this modification is to avoid the incorrect displacement of the center of each class due to an unknown fault of small dimensions with a high latency time.

When an observation k arrives, the DOFCM algorithm classifies it as outlier or as good observation taking into account the results of the training. Then, if the observation k does not belong to the outlier class, the distances between the observation and the class centers determined in the training stage are calculated. Next, the fuzzy membership degree of the observation k to each one of the c classes is obtained. The observation k will be assigned to the class where it has

the highest membership degree. The approach used in this stage by using DOFCM-KFCM without the updating of the centers of the classes is described in Algorithm 3.

Algorithm 3 Recognition (DOFCM-KFCM)

Input: data X_k , class centers \mathbf{V} , $r_{neighborhood}$, n_{max} , α , m , σ .

Output: Current State.

for $k = 1$ to $k = N$ **do**

Calculate $\eta_{neighborhood}^i$ according to Eq. (2).

Calculate $M_{neighborhood}^i$ according to Eq. (1).

if $k \notin C_{outlier}$ (according to Eq. (3)) **then**

Calculate the distances from the observation k to the center of the classes according to Eq. (11).

Calculate the membership degree of the observation k to the c good classes according to Eq. (14).

Determine to which class belongs the observation k using Eq. (16).

end if

end for

Benchmark case study: DAMADICS

Process description

In order to apply the proposed methodology to fault diagnosis in the mechanical systems the DAMADICS benchmark was selected. This benchmark represents an actuator (Bartys et al. 2006; Kourad et al. 2012) belonging to the class of intelligent electro-pneumatic devices widespread in industrial environment. The experimental data of the DAMADICS benchmark used in this paper was obtained from <http://diag.mchtr.pw.edu.pl/damadics/>. This actuator is considered as an assembly of devices consisting of:

- Control valve
- Spring-and-diaphragm pneumatic servomotor
- Positioner

The general structure of this actuator is shown in Fig. 5

The control valve acts on the flow of the fluid passing through the pipeline installation. A servomotor carries out a change in the position of the control valve plug, by acting on fluid flow rate. A spring-and-diaphragm pneumatic servomotor is a compressible fluid powered device in which the fluid acts upon the flexible diaphragm, to provide linear motion of the servomotor stem. The positioner is a device applied to eliminate the control-valve-stem miss-positions produced by the external or internal sources such as: friction, clearance in mechanical assemblies, supply pressure variations, hydrodynamic forces, among others. A description of simulated faults is shown in Table 1.

Fig. 5 Structure of benchmark actuator system

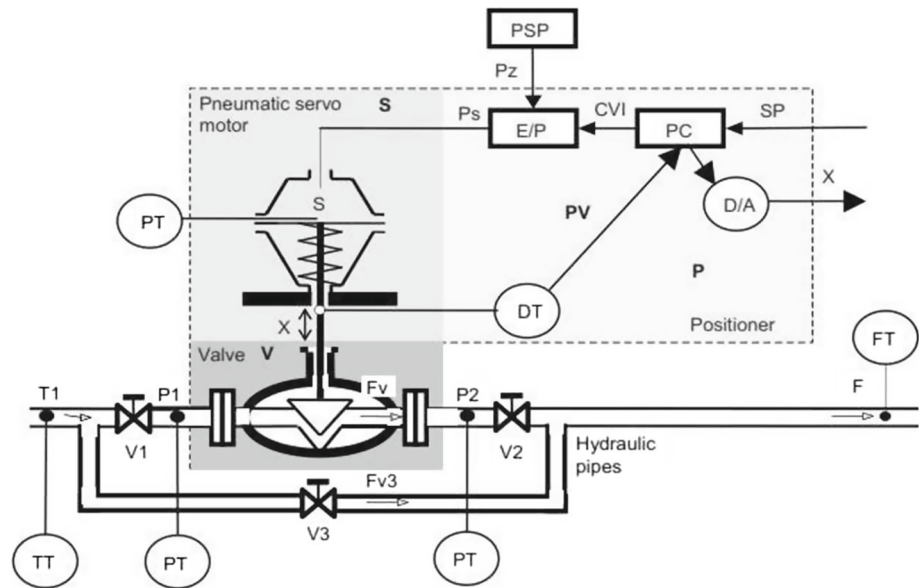


Table 1 Faults simulated in the DAMADICS

Fault	Description
1	Valve clogging
7	Critical flow
12	Electro-pneumatic transducer fault
15	Positioner spring fault
19	Flow rate sensor fault

Table 2 Measured process variables

Description	Symbol
Process control external signal	<i>CV</i>
Inlet liquid pressure	<i>P1</i>
Outlet liquid pressure	<i>P2</i>
Stem displacement	<i>X</i>
Liquid flow rate	<i>F</i>
Process value	<i>PV</i>

The set of measurements of 6 process variables shown in Table 2, were stored with a sample time of 1 second. For each one of the six process states (Normal operation and the five faults) 300 observations were stored for a total of 1800 observations. To this data set were added 300 new observations evenly distributed among the classes in order to represent the possible outliers for each class. Furthermore, white noise was added in the simulation to the measurement and process variables in order to simulate the variability presented in real world processes. The Fig. 6 shows a water level control loop in a tank with gravitational outflow.

Experiments

Two sets of three experiments each one were developed.

First, the three experiments presented in Table 3 were performed. In the first one, the step 1 (outliers determination) of the proposed classification scheme was not applied. In this experiment the KFCM algorithm was applied in the step 2. The aim of this experiment was to analyze the effect of the outliers in the final result of the classification process.

In second experiment only the DOFCM algorithm was applied in step 1. The principal aim of this experiment was to analyze the improvement in the performance of the classification process when a kernel function is introduced to obtain a better separation of the classes.

For the third experiment the DOFCM algorithm was selected to be applied in the step 1, and the KFCM algorithm was applied in the second step, respectively. The principal aim of this experiment was to analyze the results obtained in the classification process when both algorithms are adequately combined.

In these experiments the step 3 corresponding to the optimizing of the parameters of the algorithms were not applied. The values of the parameters used for the algorithms were: $Itr_max = 100, \epsilon = 10^{-5}, m = 2, \sigma = 1$.

Later, similar experiments were performed but including the step of optimizing the parameters of DOFCM and KFCM algorithms with the aim of analyzing the influence of the parameter selection (Param. Opt.) in the results of the classification process. These experiments are presented in Table 4.

It is necessary to highlight that to estimate the best parameters to be used in the DOFCM and KFCM algorithm many optimization algorithms can be used. In this paper, the results of three optimization algorithms in their typi-

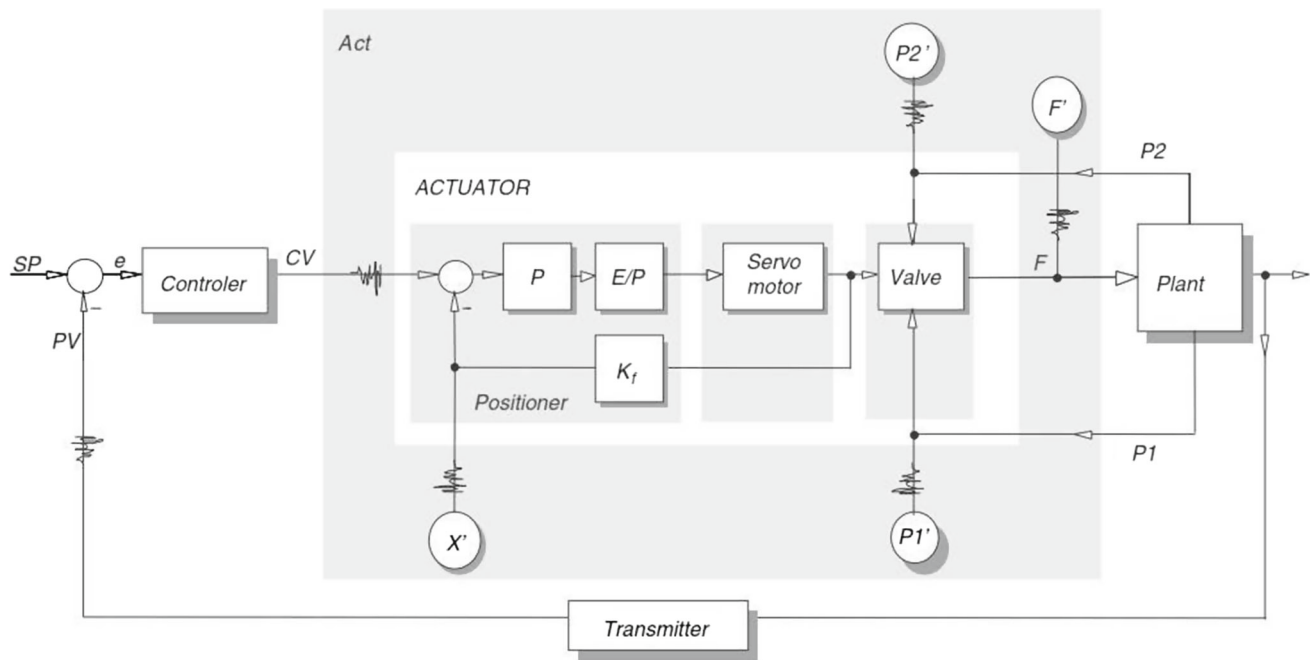


Fig. 6 Water level control loop

Table 3 Experiments performed without step 3

Experiment	Step 1	Step 2
1	–	KFCM
2	DOFCM	–
3	DOFCM	KFCM

Table 4 Experiments performed with step 3

Experiment	Step 1	Step 2	Step 3
4	–	KFCM	Param. Opt.
5	DOFCM	–	Param. Opt.
6	DOFCM	KFCM	Param. Opt.

cal structure were compared: DE (Camps Echevarría et al. 2010), ACO (Camps Echevarría et al. 2014a), and PSO (Díaz et al. 2016). The parameters used by the DE algorithm were: $C_R = 0.5$, $F_S = 0.1$, $Z = 10$ and they were obtained from (Camps Echevarría et al. 2010). In the case of ACO, the parameters selected were $k = 63$, $q_0 = 0.5$, $Z = 50$, $C_{evap} = 0.3$, $C_{inc} = 0.1$ obtained from (Camps Echevarría et al. 2014b). The parameters used by the PSO algorithm were: population size = 20, $w_{max} = 0.9$, $w_{min} = 0.4$, $c1 = 2$, $c2 = 2$ and they were obtained from (Camps Echevarría et al. 2014b).

In all cases a search space of $1 < m \leq 2$ and $0.25 \leq \sigma \leq 20$ were considered, and the following stop criteria were used:

- Criterion 1: Maximum number of evaluations of the objective function ($Eval_{max} = 100$).
- Criterion 2: ($Error = 1 - PC$) $< \epsilon = 0.00001$

DE, PSO and ACO algorithms were ran 10 times and the arithmetic mean of the parameters m , σ , and number of evaluations of the objective function ($Eval_{Fobj}$) were calculated in the experiments 4, 5 and 6. Results are shown in Table 5.

In order to determine what algorithm (DE, PSO or ACO) was better, Friedman's non-parametric statistical test was applied to the results obtained in experiments 4, 5 and 6. The result indicates that there are no significant differences in the results obtained by the 3 algorithms.

Finally, the ACO algorithm was selected, considering that it uses less evaluations of the objective function to estimate the parameters (See Table 5).

Analysis and discussion of results

Recognition stage

A very important step in the design of the fault diagnosis system is to analyze the performance of the diagnosis process. The most used criterion for this analysis is the confusion matrix (CM).

The confusion matrix allows to visualize the performance of the classifier in the classification process. Each CM_{rs} element of a confusion matrix for $r \neq s$, indicates the number

Table 5 Arithmetical mean of the parameters m , σ and $Eval_Fobj$

Experiment	Algorithm								
	DE			PSO			ACO		
	\bar{m}	$\bar{\sigma}$	$Eval_Fobj$	\bar{m}	$\bar{\sigma}$	$Eval_Fobj$	\bar{m}	$\bar{\sigma}$	$Eval_Fobj$
4	1.8615	13.716	77	1.8845	13.970	87	1.8938	14.105	76
5	1.8591	13.684	74	1.8881	13.993	83	1.8939	14.095	69
6	1.8607	13.720	83	1.8906	13.985	92	1.8953	14.154	81

Table 6 Confusion matrix for experiment 1: KFCM (NOC: 350, F1: 350, F7: 350, F12: 350, F15: 350, F19: 350)

	NOC	F1	F7	F12	F15	F19	HR (%)	E (%)
NOC	289	11	14	17	13	6	82.57	17.43
F1	16	214	12	9	2	97	61.14	38.86
F7	7	6	307	15	13	2	87.71	12.29
F12	4	13	13	301	11	8	87.71	12.29
F15	6	67	64	11	185	17	54.57	45.43
F19	88	14	17	7	8	216	61.71	38.29
GEN							72	28

Bold values indicate the main diagonal is associated with the number of observations successfully classified

of times that the classifier confuses a state r with a state s in a set of L experiments. The results obtained from the application of the proposed methodology to fault diagnosis in the modified DAMADICS data set are presented next.

The confusion matrices shown in Tables 6, 7, 8, 9, 10 and 11 were obtained using a cross validation process. Cross validation divides the dataset into complementary subsets (d), by performing the analysis on $d - 1$ subsets called the training set, and validating the analysis on the other subset called the validation set or testing set. To reduce variability d rounds of cross-validation are performed using a different partition as a validation set in each one. Finally, the validation results are averaged. Figure 7 shows the cross-validation process for four partitions of the data set. In the experiments implemented in the DAMADICS, the cross-validation was performed with 10 partitions of the data set.

Table 7 Confusion matrix for experiment 2: DOFCM (NOC: 300, F1: 300, F7: 300, F12: 300, F15: 300, F19: 300, O: 300)

	NOC	F1	F7	F12	F15	F19	O	HR (%)	E (%)
NOC	300	0	0	0	0	0	0	100	0
F1	8	211	10	6	4	60	0	70.33	29.67
F7	0	0	267	3	30	0	0	89	11
F12	0	0	1	299	0	0	0	99.67	0.33
F15	46	0	71	4	142	7	30	47.33	52.67
F19	0	0	1	0	0	299	0	99.67	0.33
O	0	0	4	0	0	0	296	98.67	1.33
GEN								86.38	13.62

Bold values indicate the main diagonal is associated with the number of observations successfully classified

Table 8 Confusion matrix for experiment 3: DOFCM-KFCM (NOC: 300, F1: 300, F7: 300, F12: 300, F15: 270, F19: 300)

	NOC	F1	F7	F12	F15	F19	HR (%)	E (%)
NOC	300	0	0	0	0	0	100	0
F1	0	257	0	0	0	43	85.67	14.33
F7	0	0	276	2	22	0	92	8
F12	0	0	0	300	0	0	100	0
F15	17	0	59	7	184	3	68.15	31.85
F19	0	0	1	0	0	299	99.67	0.33
GEN							91.30	8.70

Bold values indicate the main diagonal is associated with the number of observations successfully classified

Table 9 Confusion matrix for experiment 4: KFCM (NOC: 350, F1: 350, F7: 350, F12: 350, F15: 350, F19: 350)

	NOC	F1	F7	F12	F15	F19	HR (%)	E (%)
NOC	303	9	13	7	6	12	86.57	13.43
F1	2	300	3	6	4	35	85.71	14.29
F7	6	11	313	3	2	15	89.43	10.57
F12	7	13	6	307	2	15	87.71	12.29
F15	18	61	61	6	191	13	54.57	45.43
F19	91	12	16	2	9	220	62.86	37.14
GEN							77.81	22.19

Bold values indicate the main diagonal is associated with the number of observations successfully classified

Experiment 1

Table 6 shows the confusion matrix for experiment 1 where the operation states NOC: Normal Operation Condition, and

Table 10 Confusion matrix for experiment 5: DOFCM (NOC: 300, F1: 300, F7: 300, F12: 300, F15: 300, F19: 300, O: 300)

	NOC	F1	F7	F12	F15	F19	O	HR (%)	E (%)
NOC	300	0	0	0	0	0	0	100	0
F1	6	215	9	11	5	54	0	71.67	28.33
F7	0	0	300	0	0	0	0	100	0
F12	0	0	1	299	0	0	0	99.67	0.33
F15	55	0	52	4	145	20	24	48.33	51.67
F19	0	0	1	0	0	299	0	99.67	0.33
O	0	0	0	0	0	0	300	100	0
GEN								88.48	11.52

Bold values indicate the main diagonal is associated with the number of observations successfully classified

Table 11 Confusion matrix for experiment 6: DOFCM-KFCM (NOC: 300, F1: 300, F7: 300, F12: 300, F15: 276, F19: 300)

	NOC	F1	F7	F12	F15	F19	HR (%)	E (%)
NOC	300	0	0	0	0	0	100	0
F1	0	269	0	0	0	31	89.67	10.33
F7	0	0	300	0	0	0	100	0
F12	0	0	0	300	0	0	100	0
F15	14	0	9	5	241	7	87.32	12.68
F19	0	0	1	0	0	299	99.67	0.33
GEN							96.23	3.77

Bold values indicate the main diagonal is associated with the number of observations successfully classified

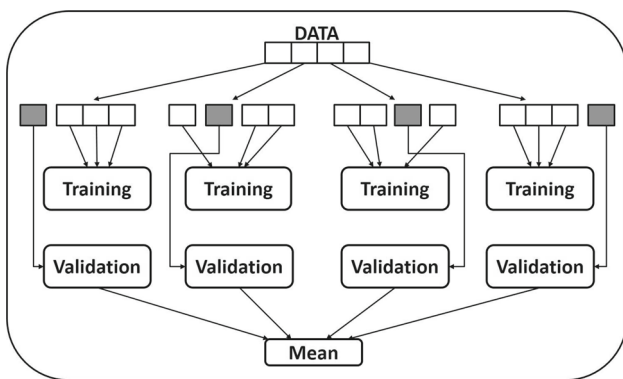


Fig. 7 Cross validation process

the faults F1, F7, F12, F15 and F19 were considered. The main diagonal is associated with the number of observations successfully classified. Since the total number of observations per class is known, the accuracy or hit rate (HR), and the overall error (E) can also be computed. The last row shows the general value of the hit rate and the error (GEN).

The results indicate the difficulty of the KFCM algorithm to obtain satisfactory classification results in the presence of outliers. This problem affects the correct classification of the different operating states, principally of F1, F15 and F19.

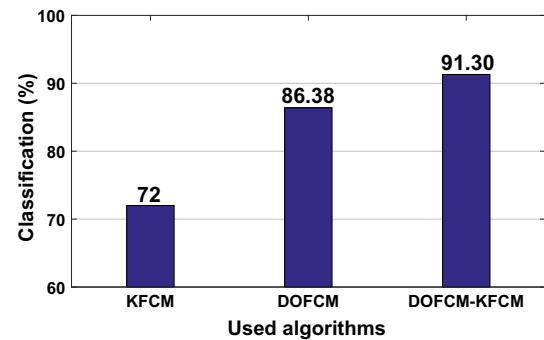


Fig. 8 Global classification (%) obtained for the experiments 1–3

Experiment 2

Table 7 shows that the DOFCM algorithm classifies as outliers 296 observations of the 300 observations added to the dataset (O class) by achieving a 98.67% of accuracy in this classification part. However, although the DOFCM algorithm identifies the outliers correctly, it is not able to obtain good results in the final classification due to overlaps between classes. This is the case of faults F1 and F15.

Experiment 3

Step 1

The classification results of the step 1 in this experiment are similar to those of the experiment 2 (Table 7). The 300 observations added as outliers were correctly identified in a 98,67%. Because 30 observations of F15 were classified in class O, the class F15 is composed of 270 observations that might be used in the next step.

Step 2

Table 8 shows the confusion matrix where the best classification results are achieved. These results are due to the eliminations of the outliers in a first step, and the application of the kernel function in the second step which allow to achieve a greater separability of the classes.

The satisfactory outcomes obtained in this experiment confirm the validity of the new methodology to fault diagno-

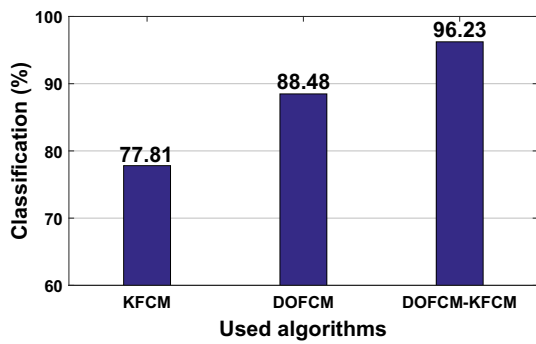


Fig. 9 Global classification (%) obtained for the experiments 4–6

sis proposed in this paper using computational intelligence techniques.

A summary of the global classification percentages obtained for each experiment are shown in Fig. 8.

Experiment 4

Table 9 shows the confusion matrix for experiment 4. As in experiment 1, the results indicate the difficulty of the KFCM algorithm to obtain satisfactory results in the classification in the presence of outliers. However, with the use of the optimized parameters m and σ , the results of the operating state classification are improved.

Experiment 5

Table 10 shows a similar behavior of the DOFCM algorithm compared with experiment 2. However, with the use of the optimized parameters m and σ , the classification results are better.

Experiment 6

Step 1

The classification results of the step 1 in this experiment are similar to those of the experiment 5 (Table 10). The 300 observations added as outliers were correctly identified. Because 24 observations of F15 were classified in class O, the class F15 is composed of 276 observations that might be used in the next step.

Step 2

Table 11 shows the results after applying the KFCM algorithm. The behavior of the DOFCM and KFCM algorithms is similar compared with experiment 3. However, with the use of the optimized parameters m and σ , the classification results are better.

The excellent results obtained in this experiment confirm the validity of the new methodology to fault diagnosis proposed in this paper using computational intelligence techniques.

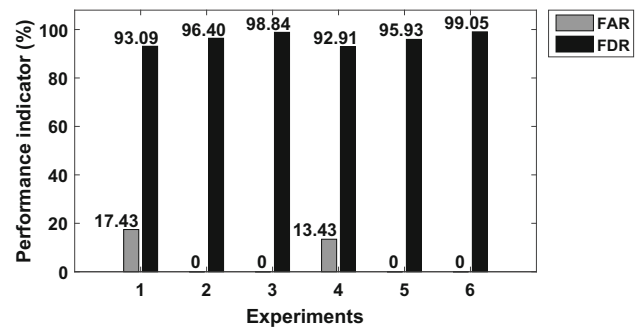


Fig. 10 Performance indicator (%) obtained for each experiment

A summary of the global classification percentages obtained for the experiments 4, 5, and 6 are shown in Fig. 9.

Comparing the results obtained in the first set of experiments with the results of the second set, it is evident the importance of selecting the best parameters for the DOFCM and KFCM algorithms which support the necessity of the stage 3 in the training process

Analysis of the number of false and missing alarms

In order to evaluate the quality of the fault detection process, the number of false and missing alarms are usually analyzed. According to Yin et al. (2012), these indicators called False Alarm Rate (FAR) and Fault Detection Rate (FDR) can be calculated by:

$$FAR = \frac{\text{No. of samples } (J > J_{lim} | f = 0)}{\text{total samples } (f = 0)} \tag{18}$$

$$FDR = \frac{\text{No. of samples } (J > J_{lim} | f \neq 0)}{\text{total samples } (f \neq 0)} \tag{19}$$

where J is the output for the used discriminative algorithms by considering the fault detection stage as a binary classification process, and J_{lim} is the threshold that determines whether one sample is classified as a fault or normal operation. The performance of a diagnosis system is satisfactory if low values of the FAR indicator and high values of the FDR indicator are obtained.

The results obtained for each experiment are summarized in Fig. 10, where it can be seen that in general form all variants have satisfactory performances. The best results are obtained in the experiment 6 which correspond with the application of the fault diagnosis methodology proposed in this paper.

Comparison with other fuzzy clustering algorithms

Recently, some fuzzy clustering algorithms with excellent results have been proposed with the aim of improving the

classification in different applications. A comparison with some of these algorithms is developed in this section.

FC-PFS algorithm

Based on theory of picture fuzzy set (PFS), [Thong and Son \(2016a\)](#) proposed a picture fuzzy model for clustering problem called FC-PFS, which was proven to get better clustering quality than other relevant methods. Based on the FCM algorithm, the picture fuzzy clustering algorithm (FC-PFS) modifies the objective function to adapt the fuzzy clustering on PFS ([Thong and Son 2016a](#)). The modification includes two points. The first one inherits from FCM’s objective function where the membership degree μ are replaced by $\mu(2 - \xi)$ which means that one data element belonging to a cluster has both: high value of positive degree and low value of refusal degree ([Thong and Son 2016a](#)). The second point is to add the entropy information to the objective function which helps the algorithm to reduce the neutral and refusal degree of an element to become a member of the cluster. The entropy information plays an important role to enhance the clustering quality ([Thong and Son 2016a](#)). The values of the parameters used are: $Itr_max = 100$, $\epsilon = 10^{-5}$, $m = 2$ and $\alpha = 0.6$ (where $\alpha \in (0, 1]$ is an exponent coefficient used to control the refusal degree in PFS sets).

PFCA-CD algorithm

In the paper [Thong and Son \(2016b\)](#), is proposed a novel picture fuzzy clustering algorithm for complex data called PFCA-CD that deals with both mix data type and distinct data structures. The idea of this method is the modification of FC-PFS, using a measurement for categorical attributes, multiple centers of one cluster and an evolutionary strategy - Particle Swarm Optimization (PSO). Therein, the multiple centers are used to deal with complex structure of data because data with complex structures have many different shapes that cannot be represented by one center. The values of the parameters used are: $Itr_max = 100$, $\epsilon = 10^{-5}$, $m = 2$, $\alpha = 0.6$, $C_1 = C_2 = 1$ (where $C_1, C_2 \geq 0$ are PSO’s parameters. Generally, C_1, C_2 are often set as 1).

DBWFCM algorithm

A new fuzzy clustering method called density-based weighted FCM (DBWFCM) is proposed in [Li et al. \(2016\)](#). In this algorithm, the weight of an object is decided by the density of the objects around this object. To more objects around this object, the weight of this object is bigger. That means the object that has bigger weight is more likely to be a cluster center. There are two stages of the density-based weighted FCM. The first stage is designed to calculate the weights of every object, the second stage is the clustering stage. The val-

ues of the parameters used are: $Itr_max = 100$, $\epsilon = 10^{-5}$ and $m = 2$.

Results of the comparison

To establish the comparison, the same data set used in the last experiments was used. Tables 12, 13 and 14 show the results of the confusion matrices related with the algorithms used in the comparison. As it can be observed, the results indicate

Table 12 Confusion matrix: FC-PFS (NOC: 350, F1: 350, F7: 350, F12: 350, F15: 350, F19: 350)

	NOC	F1	F7	F12	F15	F19	HR (%)	E (%)
NOC	299	13	14	8	4	12	85.43	14.57
F1	7	293	3	6	10	31	83.71	16.29
F7	8	11	297	9	5	20	84.86	15.14
F12	11	9	8	310	2	10	88.57	11.43
F15	16	29	37	14	235	19	67.14	32.86
F19	51	12	10	6	9	262	74.86	25.14
GEN							80.76	19.24

Bold values indicate the main diagonal is associated with the number of observations successfully classified

Table 13 Confusion matrix: PFCA-CD (NOC: 350, F1: 350, F7: 350, F12: 350, F15: 350, F19: 350)

	NOC	F1	F7	F12	F15	F19	HR (%)	E (%)
NOC	305	10	11	9	5	10	87.14	12.86
F1	5	298	2	8	5	32	85.14	14.86
F7	9	10	301	8	5	17	86.00	14.00
F12	6	10	8	312	3	11	89.14	10.86
F15	12	31	33	9	248	17	70.86	29.14
F19	43	9	11	3	8	276	78.86	21.14
GEN							82.86	17.14

Bold values indicate the main diagonal is associated with the number of observations successfully classified

Table 14 Confusion matrix: DBWFCM (NOC: 350, F1: 350, F7: 350, F12: 350, F15: 350, F19: 350)

	NOC	F1	F7	F12	F15	F19	HR (%)	E (%)
NOC	301	9	13	7	6	14	86.00	14.00
F1	4	297	1	12	7	29	84.86	15.14
F7	7	13	302	10	4	14	86.29	13.71
F12	8	7	11	308	7	9	88.00	12.00
F15	14	26	35	13	244	18	69.71	30.29
F19	46	11	9	5	10	269	76.86	23.14
GEN							81.95	18.05

Bold values indicate the main diagonal is associated with the number of observations successfully classified

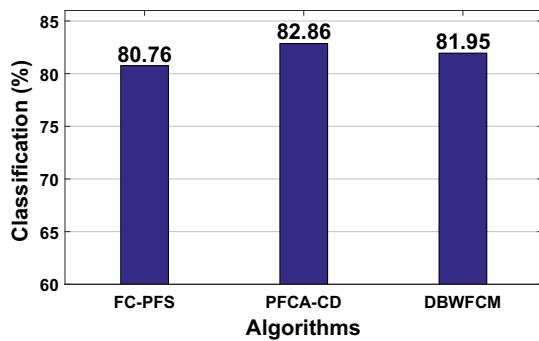


Fig. 11 Global classification (%) obtained for each algorithm

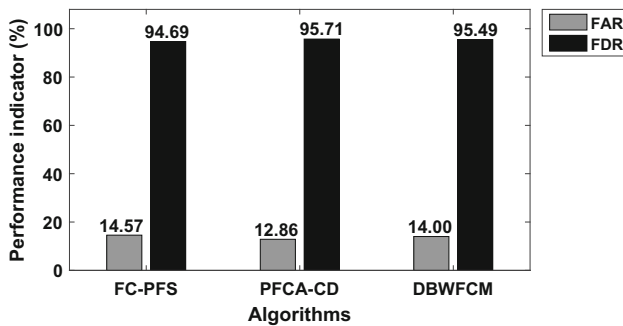


Fig. 12 Performance indicator (%) obtained for each algorithm

the difficulty of these algorithms to obtain satisfactory results in the classification in the presence of outliers.

A summary of these results can be seen in Fig. 11, where the global classification percentages obtained for each algorithm are showed. The results obtained with the algorithms used in the comparison (FC-PFS, PFCA-CD, DBWFCM) are smaller than the result obtained with the proposal made in this paper (96.23%).

Figure 12 shows the FAR and FDR values obtained for each one of these algorithms. It is evident that the results obtained with the proposal made in this paper (FAR = 0% and FDR = 99.05%) constitute the best outcomes but the best practice is to support this appreciation by applying statistical tests (García and Herrera 2008; García et al. 2009; Luengo et al. 2009).

Statistical tests

First, the non-parametric Friedman test is applied in order to demonstrate that there is at least one algorithm whose results have significant differences with respect to results of the others. Afterwards, if the null-hypothesis of the Friedman test is rejected, it is necessary to make a comparison in pairs to determine the best algorithm(s). For this, the non-parametric Wilcoxon test is applied.

Table 15 Results of the Wilcoxon test

	1 vs 2	1 vs 3	1 vs 4	2 vs 3	2 vs 4	3 vs 4
$\sum R^+$	0	0	0	54	0	0
$\sum R^-$	55	55	55	1	55	55
T	0	0	0	1	0	0
$T_{\alpha=0.05}$	8	8	8	8	8	8
Winner	2	3	4	2	4	4

Table 16 Final result of the comparison between algorithms

Algorithm	No. Wins	Ranking
FC-PFS	0	4
PFCA-CD	2	2
DBWFCM	1	3
DOFCM-KFCM	3	1

Friedman test

In this case, for four algorithms ($k = 4$) and 10 data sets due to the cross validation ($N = 10$), the value obtained to the Friedman statistic was $F_F = 241$. With $k = 4$ and $N = 10$, F_F is distributed according to the F distribution with $4 - 1 = 3$ and $(4 - 1) \times (10 - 1) = 27$ degrees of freedom. The critical value of $F(3,27)$ for a level of significance $\alpha = 0.05$ is 2.9604, then the null-hypothesis is rejected ($F(3, 27) < F_F$). This means that at least, there is an algorithm whose results differ significantly from the rest.

Wilcoxon test

Table 15 shows the results of the comparison in pairs of the algorithms (1: FC-PFS, 2: PFCA-CD, 3: DBWFCM, 4: DOFCM-KFCM) using the Wilcoxon test. The first two rows contain the values of the sum of the positive (R^+) and negative (R^-) rank for each comparison established. The next two rows show the statistical values T and the critical value of T for a level of significance $\alpha = 0.05$. The last row indicates which algorithm was the winner in each comparison. The summary in Table 16 shows how many times each algorithm was the winner. This results validates once again the methodology proposed in this paper.

Conclusions

The mechanical systems are fundamental elements in the manufacturing industry and a large number of faults in these industries occur in such systems. In the present paper a new robust scheme to fault diagnosis in mechanical systems using computational intelligence techniques is proposed to

decrease the unfavorable impact of these faults in the productivity, the environment and the safety of the operators.

In this paper, several experiments were presented with the aim of demonstrating that the best strategy is the one that integrates the DOFCM, KFCM algorithms and an optimization algorithm. In this paper the ACO algorithm was used. The DOFCM algorithm was used in the first step for preprocessing the data to remove the outliers. The KFCM algorithm was used in the second step for the data classification to make use of the advantages introduced by the kernel function in the separability of the classes, in order to obtain better classification outcomes. Finally, the ACO algorithm was used to optimize the parameters of the algorithms used in the previous steps.

Three fuzzy clustering algorithms recently presented in the scientific literature were selected to establish a comparison with the proposed methodology. After comparing the obtained results, it was demonstrated that the proposed methodology obtained the best performance.

Acknowledgements The authors acknowledge the financial support provided by FAPERJ, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro; CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico; CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, research supporting agencies from Brazil; UERJ, Universidade do Estado do Rio de Janeiro and CUJAE, Universidad Tecnológica de La Habana José Antonio Echeverría.

References

- Aydin, I., Karakose, M., & Akin, E. (2014). An approach for automated fault diagnosis based on a fuzzy decision tree and boundary analysis of a reconstructed phase space. *ISA Transactions*, 53, 220–229.
- Bartys, M., Patton, R., Syfert, M., de las Heras, S., & Quevedo, J. (2006). Introduction to the DAMADICS actuator FDI benchmark study. *Journal of Control Engineering Practice*, 14, 577–596.
- Bedoya, C., Uribe, C., & Isaza, C. (2012). *Unsupervised feature selection based on fuzzy clustering for fault detection of the Tennessee Eastman process. Advances in artificial intelligence (LNAI 7637 pp. 350 – 360)*, Springer-Verlag.
- Bernal de Lázaro, J. M., Llanes-Santiago, O., Prieto Moreno, A., Knupp, D. C., & Silva-Neto, A. J. (2016). Enhanced dynamic approach to improve the detection of small-magnitude faults. *Chemical Engineering Science*, 146, 166–179.
- Bernal de Lázaro, J. M., Prieto Moreno, A., Llanes-Santiago, O., & Silva Neto, A. J. (2015). Optimizing kernel methods to reduce dimensionality in fault diagnosis of industrial systems. *Computers & Industrial Engineering*, 87, 140–149.
- Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum.
- Bocaniala, C. D., Costa, J. S. D., & Palade, V. (2005). Fuzzy-based refinement of the fault diagnosis task in industrial devices. *Journal of Intelligent Manufacturing*, 16(6), 599–614. doi:10.1007/s10845-005-4365-z.
- Botia, J., Isaza, C., Kempowsky, T., LeLann, M. V., & Aguilar-Martín, J. (2013). Automation based on fuzzy clustering methods for monitoring industrial processes. *Engineering Applications of Artificial Intelligence*, 26, 1211–1220.
- Camps Echevarría, L., de Campos Velho, H. F., Becceneri, J. C., Silva Neto, A. J., & Llanes-Santiago, O. (2014a). The fault diagnosis inverse problem with ant colony optimization and ant colony optimization with dispersion. *Applied Mathematics and Computation*, 227, 687–700.
- Camps Echevarría, L., Llanes-Santiago, O., Hernández Fajardo, J. A., Silva Neto, A. J., & Jiménez Sánchez, D. (2014b). A variant of the particle swarm optimization for the improvement of fault diagnosis in industrial systems via faults estimation. *Engineering Applications of Artificial Intelligence*, 28, 36–51.
- Camps Echevarría, L., Llanes-Santiago, O., & Silva Neto, A. J. (2010). Chap. An approach for Fault diagnosis based on bio-inspired strategies. In: *Nature inspired cooperative strategies for optimization (NICSO 2010)* (Vol. 284, pp. 53 – 63), Springer.
- Chintalapudi, K. K., & Kam, M. (1998). A noise resistant fuzzy c-means algorithm for clustering. *IEEE Conference on Fuzzy Systems Proceedings*, 2, 1458–1463.
- Choudhary, A. K., Harding, J. A., & Tiwari, M. K. (2008). Data mining in manufacturing: A review based on the kind of knowledge. *Journal of Intelligent Manufacturing*, 20(5), 501–521.
- da Silva, A. M., Povinelli, R. J., & Demerdash, N. A. (2008). Induction machine broken bar and stator short-circuit fault diagnostics based on three-phase stator current envelopes. *IEEE Transactions on Industrial Electronics*, 55, 1310–1318.
- Dave, R. N. (1991). Characterization and detection of noise in clustering. *Pattern Recognition Letters*, 12, 657–664.
- Dave, R. N., & Krishnapuram, R. (1997). Robust clustering methods: A unified view. *IEEE Transactions on Fuzzy Systems*, 5, 270–293.
- Díaz, C. A., Echevarría, L. C., Prieto-Moreno, A., Neto, A. J. S., & Llanes-Santiago, O. (2016). A model-based fault diagnosis in a nonlinear bioreactor using an inverse problem approach and evolutionary algorithms. *Chemical Engineering Research and Design*, 114, 18–29.
- Ding, S. X. (2008). *Model-based fault diagnosis techniques*. Berlin: Springer.
- Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the 2nd ACM SIGKDD* (pp. 226 – 231), Portland, Oregon.
- Fan, J., & Wang, Y. (2014). Fault detection and diagnosis of non-linear non-Gaussian dynamic processes using kernel dynamic independent component analysis. *Information Sciences*, 259, 369–379.
- García, S., & Herrera, F. (2008). An extension on statistical comparisons of classifiers over multiple data sets for all pairwise comparisons. *Journal of Machine Learning Research*, 9, 2677–2694.
- García, S., Molina, D., Lozano, M., & Herrera, F. (2009). A study on the use of non-parametric tests for analyzing the evolutionary algorithms behaviour: A case study on the CEC'2005 special session on real parameter optimization. *Journal of Heuristic*, 15, 617–644.
- Gosain, A., & Dahika, S. (2016). Performance analysis of various fuzzy clustering algorithms: A review. In: *7th international conference on communication, computing and virtualization* (Vol. 79, pp. 100 – 111).
- Hou, T. H. T., Liu, W. L., & Lin, L. (2003). Intelligent remote monitoring and diagnosis of manufacturing processes using an integrated approach of neural networks and rough sets. *Journal of Intelligent Manufacturing*, 14(2), 239–253. doi:10.1023/A:1022911715996.
- Hu, Q., He, Z., Zhang, Z., & Zi, Y. (2007). Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble. *Mechanical Systems and Signal Processing*, 21(2), 688–705.
- Hwang, I., Kim, S., Kim, Y., & Seah, C. (2010). A survey of fault detection, isolation, and reconfiguration methods. *IEEE Transactions on Control Systems Technology*, 18, 636–656.

- Isermann, R. (2011). *Fault-diagnosis applications: Model-based condition monitoring: Actuators, drives, machinery, plants, sensors and fault-tolerant systems*. Berlin: Springer.
- Jahromi, A. T., Er, M. J., Li, X., & Lim, B. S. (2016). Sequential fuzzy clustering based dynamic fuzzy neural network for fault diagnosis and prognosis. *Neurocomputing*, *196*, 31–41.
- Jiang, X. L., Wang, Q., He, B., Chen, S. J., & Li, B. L. (2016). Robust level set image segmentation algorithm using local correntropy-based fuzzy c-means clustering with spatial constraints. *Neurocomputing*, *207*, 22–35.
- Karami, F., Poshtan, J., & Poshtan, M. (2010). Detection of broken rotor bars in induction motors using nonlinear Kalman filters. *ISA Transactions*, *49*, 189–195.
- Kaur, P. (2011). A density oriented fuzzy c-means clustering algorithm for recognising original cluster shapes from noisy data. *International Journal of Innovative Computing and Applications*, *3*, 77–87.
- Kaur, P., Soni, A., & Gosain, A. (2013). Robust kernelized approach to clustering by incorporating new distance measure. *Engineering Applications of Artificial Intelligence*, *26*, 833–847.
- Kesemen, O., Tezel, O., & Ozkul, E. (2016). Fuzzy c-means clustering algorithm for directional data (FCM4DD). *Expert Systems with Applications*, *58*, 76–82.
- Kourd, Y., Lefebvre, D., & Guersi, N. (2012). FDI with neural network models of faulty behaviours and fault probability evaluation: Application to DAMADICS. In: *8th IFAC symposium on fault detection, supervision and safety of technical processes (SAFE-PROCESS)* (pp. 744 – 749), August 29–31, 2012.
- Krishnapuram, R., & Keller, J. M. (1993). A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems*, *1*, 98–110.
- Leski, J. M. (2016). Fuzzy c-ordered-means clustering. *Fuzzy Sets and Systems*, *286*, 114–133.
- Liu, Z., He, Z., Guo, W., & Tang, Z. (2016). A hybrid fault diagnosis method based on second generation wavelet de-noising and local mean decomposition for rotating machinery. *ISA Transactions*, *61*, 211–220.
- Liu, Q., & Lv, W. (2009). The study of fault diagnosis based on particle swarm optimization algorithm. *Computer and Information Science*, *2*, 87–91.
- Li, Y., Yang, G., He, H., Jiao, L., & Shang, R. (2016). A study of large-scale data clustering based on fuzzy clustering. *Soft Computing*, *20*, 3231–3242.
- Li, C., Zhou, J., Kou, P., & Xiao, J. (2012). A novel chaotic particle swarm optimization based fuzzy clustering algorithm. *Neurocomputing*, *83*, 98–109.
- Lobato, F., Steffen, F., Jr., & Silva Neto, A. (2009). Solution of inverse radiative transfer problems in two-layer participating media with differential evolution. *Inverse Problems in Science and Engineering*, *18*, 183–195.
- Luengo, J., García, S., & Herrera, F. (2009). A study on the use of statistical tests for experimentation with neural networks: Analysis of parametric test conditions and non-parametric tests. *Expert Systems with Applications*, *36*, 7798–7808.
- Pakhira, M., Bandyopadhyay, S., & Maulik, S. (2004). Validity index for crisp and fuzzy clusters. *Pattern Recognition*, *37*, 487–501.
- Pal, N. R., Pal, K., Keller, J. M., & Bezdek, J. C. (2005). A possibilistic fuzzy c-means clustering algorithm. *IEEE Transactions on Fuzzy Systems*, *13*, 517–530.
- Pang, Y. Y., Zhu, H. P., & Liu, F. M. (2014). Fault diagnosis method based on kpca and selective neural network ensemble. *Advanced Materials Research*, *915*, 1272–1276.
- Patan, K. (2008). *Artificial neural networks for the modelling and fault diagnosis of technical processes*. Berlin: Springer.
- Rodríguez Ramos, A., Domínguez Acosta, C., Rivera Torres, P. J., Serrano Mercado, E. I., Beauchamp Baez, G., Rifón, L. A., et al. (2016). An approach to multiple fault diagnosis using fuzzy logic. *Journal of Intelligent Manufacturing*. doi:10.1007/s10845-016-1256-4.
- Salto, R., & Weber, R. (2016). A rough-fuzzy approach for support vector clustering. *Information Sciences*, *339*, 353–368.
- Seera, M., Lim, C. P., Loo, C. K., & Singh, H. (2015). A modified fuzzy min–max neural network for data clustering and its application to power quality monitoring. *Applied Soft Computing*, *28*, 19–29.
- Sharifi, R., & Ebrahimi, M. (2011). Detection of stator winding faults in induction motors using three-phase current monitoring. *ISA Transactions*, *50*, 14–20.
- Sina, S., Sadough, Z. N., & Khorasani, K. (2014). Dynamic neural network-based fault diagnosis of gas turbine engines. *Neurocomputing*, *125*, 153–165.
- Thong, P. H., & Son, L. H. (2016). Picture fuzzy clustering: A new computational intelligence method. *Soft Computing*, *20*, 3549–3562.
- Thong, P. H., & Son, L. H. (2016). Picture fuzzy clustering for complex data. *Engineering Applications of Artificial Intelligence*, *56*, 121–130.
- Venkatasubramanian, V., Rengaswamy, R., & Kavuri, S. N. (2003a). A review of process fault detection and diagnosis, part 1: Quantitative model-based methods. *Computers and Chemical Engineering*, *27*, 293–311.
- Venkatasubramanian, V., Rengaswamy, R., & Kavuri, S. N. (2003b). A review of process fault detection and diagnosis, part 2: Qualitative models and search strategies. *Computers and Chemical Engineering*, *27*, 313–326.
- Vonga, C. M., Wong, P. K., & Wong, K. I. (2014). Simultaneous-fault detection based on qualitative symptom descriptions for automotive engine diagnosis. *Applied Soft Computing*, *22*, 238–248.
- Wang, J., & Hu, H. (2009). Vibration-based fault diagnosis of pump using fuzzy technique. *Measurement*, *39*, 176–185.
- Wu, K., & Yang, M. (2005). A cluster validity index for fuzzy clustering. *Pattern Recognition*, *26*, 1275–1291.
- Xu, Z., Li, Y., Wang, Z., & Xuan, J. (2016). A selective fuzzy artmap ensemble and its application to the fault diagnosis of rolling element bearing. *Neurocomputing*, *19*, 25–35.
- Yin, S., Ding, S. X., Haghani, A., Hao, H., & Zhang, P. (2012). A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. *Journal of Process Control*, *22*(9), 1567–1581.
- Zhang, D. Q., & Chen, S. C. (2003). Clustering incomplete data using kernel based fuzzy c-means algorithm. *Neural Process Letters*, *18*, 155–162.
- Zhang, D. Q., & Chen, S. C. (2004). A novel kernelized fuzzy c-means algorithm with application in medical image segmentation. *Artificial Intelligence in Medicine*, *32*, 37–50.
- Zhang, L., Lu, W., Liu, X., Pedrycz, W., & Zhong, C. (2016). Fuzzy c-means clustering of incomplete data based on probabilistic information granules of missing values. *Knowledge-Based Systems*, *99*, 51–70.