Observer biased clustering in wind turbine fault diagnosis

Bernardo Alpoim Filipe Ferreira Pimenta
bernardo.pimenta@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

December 2021

Abstract

Over the last two decades wind power has undergone an exponential growth globally. As the demand for wind power systems increases, efficiency is sought to be maximized and operation and maintenance costs reduced. Condition monitoring (CM) systems have become a field of high interest within the wind turbine world. The main components of wind turbines constitute the focus of CM, as they are responsible for frequent, large repair costs and operational downtime. Within the main components, the gearbox accounts for one of the highest failure rates and is the component that causes the greatest amount of downtime. This work will focus on the application of fuzzy clustering for wind turbine gearboxes fault detection and classification. The Gath-Geva clustering algorithm is explored by applying the observer biased clustering framework. The notion of an observer allows clustering to be an interactive process, providing an intuitive way to control cluster formation and enabling domain knowledge to be incorporated in the process. A domain expert can choose the level of granularity and is able to select a particular region of the data space for a detailed view. The Gath-Geva with Focal Point algorithm is tested with wind turbine gearbox vibrational data and compared with its unbiased version, the Fuzzy C-Means biased and unbiased algorithms. Two metrics are employed to validate internal clustering: the well known Xie-Beni index and the Kim-Lee index, the latter of which is based on relative degree of sharing. The algorithms are compared by performing several independent runs and using the distribution of the Adjusted Rand Index external validation metric.
Keywords: Fuzzy clustering, Wind turbine fault detection, Observer biased clustering, Focal Point, Gath Geva with Focal Point (GGFP)

1. Introduction

From all the components within a wind turbine drive train, the gearbox is responsible for 12% of all failures (the second highest rate) [21] and is also the component whose failure causes the largest downtime [8]. These factors make it essential to ensure the healthy and stable operation of this type of equipment, to address this problem, gearbox fault diagnosis has already been the focus of many researchers in the condition monitoring field.

Condition Monitoring (CM) aims to use measured data to predict deterioration and failure of machine components, which has led CM systems to become highly sought after in this industry. A review of different techniques used in CM of wind turbines (WT) can be consulted in [15], this review includes traditional techniques such as vibration signals, acoustic emission, or ultrasonic testing, it has been proven that vibrational analysis is the technique that gives more information about faults in rotating machinery, making vibration sensors widely used in wind turbine applications [25]. The survey [26] gathers an extensive list of approaches based on vibration based condition monitoring for wind turbine gearboxes, ranging from signal processing methods to fault detection methods including Artificial Intelligence and Machine Learning based approaches.

Methods based on supervised learning can be used for fault diagnosis after having been trained with known fault training samples. However, only the patterns that are found in the training samples can be classified. Wrong diagnosis can often occur when dealing with unknown faults, making it hard to use supervised methods effectively. Thus, the application of unsupervised pattern recognition methods, such as clustering, becomes relevant.

Fuzzy clustering methods applied to WT gearbox fault detection include the application of the K-means Clustering Method [20] and the unsupervised learning method Kernel C-Means [12]. In comparison to WT gearbox fault detection, applications that involve other machine parts, such as bearings, display a much larger list of different approaches in the literature, ranging from different variations of the Fuzzy C-Means algorithm, to the application of the Gustafson-Kessel and Gath-Geva algorithm [13, 14, 19, 27]. Motivated by the success of the application of observer biased fuzzy clustering in fault detection and classification for bearing CM, the Gath-Geva with Focal Point (GGFP) clustering algorithm is here applied to untested, real world data and compared with other state of the art fuzzy clustering algorithms.

2. Fuzzy Clustering

In general terms, clustering can be divided into two major groups: hard clustering and soft clustering, the latter of which includes fuzzy clustering. Within each category there is a diverse number of algorithms, utilizing different approaches for the way partitions are formed.

In hard, or non-fuzzy, clustering data is assigned to clusters such that the degree of membership of each data point to a particular cluster is either 0 or 1. These types of clusters are called crisp clusters. In other words, a given data point belongs to exactly a single cluster. There is an abundance of hard clustering algorithms, among which some of the most well-known are the K-Means algorithm and hierarchical clustering [23].

Utilizing notions of fuzzy sets, data points in soft clustering may belong to more than one cluster. Each data point is then attributed membership values which indicate the likelihood of belonging to different clusters [2]. In this work, the focus is directed to the Gath-Geva algorithm and its observer biased variant.

2.1. Gath-Geva

The Gath-Geva (GG) clustering algorithm was originally proposed by Gath and Geva in [7], and is also known as fuzzy maximum-likelihood clustering. The GG algorithm detects hyper-ellipsoidal clusters with different orientations, sizes, and densities; it aims at minimizing the objective function:

$$J = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{m} ||x_j - v_i||^2 \tag{1}$$

where the distance metric is given by:

$$||\boldsymbol{x}_j - \boldsymbol{v}_i||^2 = \frac{|\boldsymbol{F}_i|^{1/2}}{\Pr(i)} \exp[-\frac{1}{2}(\boldsymbol{x}_j - \boldsymbol{v}_i)^T \boldsymbol{F}_i^{-1} (\boldsymbol{x}_j - \boldsymbol{v}_i)] \tag{2}$$

where $\Pr(i)$ is the priori probability of the i-th cluster, i.e.,

$$\Pr(i) = \frac{\sum_{j=1}^{n} u_{ij}^{m}}{\sum_{j=1}^{n} \sum_{\iota=1}^{c} u_{\iota k}^{m}} \tag{3}$$

and $\boldsymbol{F}_i$ and $|\boldsymbol{F}_i|$ are the fuzzy covariance matrix of the i-th cluster (6) and its determinant, respectively. Once (2) is computed the updates of u_{ij}, and $\boldsymbol{v}_i$ can be found by (4) and (5), respectively.

$$u_{ij} = [\sum_{k=1}^{c} \left(\frac{||\boldsymbol{x}_j - \boldsymbol{v}_i||^2}{||\boldsymbol{x}_k - \boldsymbol{v}_i||^2} \right)^{\frac{1}{m-1}}]^{-1} \tag{4}$$

The updating expression for the prototypes:

$$\boldsymbol{v}_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} \boldsymbol{x}_j}{\sum_{j=1}^{n} u_{ij}^{m}} \tag{5}$$

$$\boldsymbol{F}_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} (\boldsymbol{x}_j - \boldsymbol{v}_i)(\boldsymbol{x}_j - \boldsymbol{v}_i)^T}{\sum_{j=1}^{n} u_{ij}^{m}} \tag{6}$$

2.2. Evaluation of clustering results

There is a wide variety of indexes in the literature that have been extensively explored [1]. In this work, two internal indexes and one external index are applied. The internal indexes are the Xie-Beni (XB) index [22] and the Kim-Lee (KL) index [10]. For the internal validity measure, as the data used for testing is labeled, the Adjusted Rand Index was the measure chosen.

2.2.1 Xie-Beni Index

Tthe XB index has been proven to provide great performance and to be quite reliable [16] when evaluating partitions produced by the FCM. The Xie Beni index focus in the identification of compact and well-separated clusters, for computational purposes the inverse of the XB index was applied and is given by:

$$\boldsymbol{XB}^{-1} = \frac{n\, min_{i \neq j} ||\boldsymbol{v}_i - \boldsymbol{v}_j||^2}{\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{m} ||\boldsymbol{x}_j - \boldsymbol{v}_i||^2} \tag{7}$$

This index can be seen as the ratio between degrees of intra-cluster distance and inter-cluster distance where the numerator represents the minimal separation between fuzzy clusters and the denominator is the sum of the compactness of each fuzzy cluster. The optimal partition is then obtained by maximizing equation 7.

2.2.2 Kim-Lee Index

The shapes and sizes of the clusters in the GG algorithm differs from the spherical and same sized clusters of the FCM. This limits the applicability of indexes, like the Xie-Beni, that rely solely on the clusters' centroid distance to evaluate cluster separation. Since the GG algorithm involves the

Gaussian metric, these type of indexes cannot differentiate two different partitions with the same centroid distance and different orientations. The KL index's goal is to calculate the average overlap between clusters using the definition of the relative degree of sharing between fuzzy clusters. The higher the membership values of a data point to a pair of clusters, the higher the relative degree of sharing between the clusters, which is an indication of cluster overlap. To take into account overlapped data points a weighing parameter is introduced and highly overlapped data points are given a bigger weight over data that are classified clearly. The relative degree of sharing between two fuzzy clusters is then defined by the weighted sum of the relative degrees of sharing at each data point x_k:

$$\boldsymbol{S(C_i, C_j)} = \sum_{k=1}^{n} [c \cdot [u_{ik} \wedge u_{jk}] h(x_k)] \tag{8}$$

where, the fuzzy AND operator [24] is employed : $u_{ik} \wedge u_{jk} = min(u_{ik}, u_{jk})$ and the entropy of data points is used as weighing parameter: $h(x_k) = -\sum_{i=1}^{c} u_{ik} log\, u_{ik}$, this allows vague (unclearly classified) points to be have a a bigger value, meaning that partitions with highly overlapped data will be penalized. This index has as its only input the partition matrix, u_{ij}, and considers a partitioning optimal when the degree of overlap between clusters is minimal. The KL index is then defined as the average relative degree of sharing for all possible cluster pairs:

$$\boldsymbol{KL} = \frac{2}{c\,(c-1)} \sum_{i=1}^{c-1} \sum_{j=i+1}^{c} \sum_{k=1}^{n} [c \cdot [u_{ik} \wedge u_{jk}] h(x_k)] \tag{9}$$

The KL index can measure the ambiguity of a certain partitioning as well as the geometrical property of overlap between clusters. It possesses the advantage over other indexes of calculating the separation between clusters without making use of inter-center distances.

2.2.3 Rand Index and Adjusted Rand Index

For external validation, cluster validity measures Rand Index (RI) and Adjusted Rand Index (ARI), [9], are employed . These indexes measure the similarity of two partitions of a data set $X = (x_1, ..., x_n)$, the ground truth partition, $\mathcal{G}$, and a hypothesis partition $\mathcal{H}$ generated by a clustering algorithm. The RI of random partitions is not constant and two random partitions can produce high RI values even if they don't represent the structure in the data. In order to deal with these drawbacks, the RI is corrected for chance resulting in the more reliable Adjusted Rand Index:

$$\mathrm{ARI}(\mathcal{G}, \mathcal{H}) = \frac{\sum_{i,j} \binom{n_{ij}}{2} - [\sum_i \binom{n_i}{2} \sum_j \binom{n_{.j}}{2} / \binom{n}{2}]}{\frac{1}{2}[\sum_i \binom{n_i}{2} + \sum_j \binom{n_j}{2}] - [\sum_i \binom{n_i}{2} \sum_j \binom{n_{.j}}{2} / \binom{n}{2}]} \tag{10}$$

where n_{ij} is the number of elements that cluster together in the subsets $\mathcal{G}_i$ and $\mathcal{H}_j$; n_i,n_j are the number of elements in $\mathcal{G}_i$ and $\mathcal{H}_j$, respectively. The ARI also attains its maximum value at 1 when both partitions completely agree.

3. Observer Biased Clustering

The observer biased variant applied to fuzzy clustering was first introduced in [4]. In theory, this approach can be applied to any clustering algorithm that seeks to optimize an objective function. The purpose of the observer biased technique is to give a clustering algorithm the ability to change the position from which the data is observed, and thus allowing for the possibility of viewing different perspectives of the data.

The inspiration for this algorithm surged from a metaphor found in daily life depicted in [5]. It uses the analogy of human perception of objects depending on the point of observation;v the closer the location of the observer, the more distinct each object becomes. Conversely, the further away the observer stands from the objects, the less detail is visualized and the group objects begins to merge into a single element. This is akin to the effect of the zoom property found in an optical lens which can cover a wide array of perspectives, ranging from fine details when zoomed in and the "bigger picture" when zoomed out.

The approach follows a statistics concept known as shrinkage. Usually, shrinkage is applied with the goal of improving an unbiased estimate by adding a regularization term. Clustering algorithms based on shrinkage techniques have been proposed and have proven through experimental results to have advantages in comparison to unbiased traditional algorithms [6, 17]. Although the observer biased approach is based on shrinkage, it differs from previously proposed algorithms due to the regularization coefficient being different from the ones previously suggested, seeing that it originated from a distinct motivation.

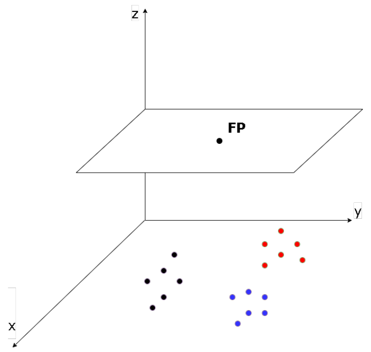


Figure 1: Data set in $\mathbb{R}^2$ focal point in a higher dimension, $\mathbb{R}^3$

The placement of the observer becomes a matter of significance, a pertinent question is to ask where should the observer be located. The location of the observer is not constrained, it can be placed anywhere in the data space or even in a different dimensional space. To fulfill the previously mentioned metaphor, the observer is placed in a higher dimensional space than the data, this also has the benefit of easily being able to distinguish the observer point from data points. Figure 1 illustrates this, in the case where the data belongs to the two dimensional space, $\mathbb{R}^2$, and focal point is placed in a higher dimension of the data, the tri-dimensional space $\mathbb{R}^3$.

In summary, this approach allows for an intuitive way to control the process of cluster formation, by incorporating the knowledge of domain experts which can select the most appropriate level of granularity and the relevance of regions of the data space.

4. Gath Geva with a Focal Point

The GG objective function is modified to include an additional term which depends on $\mathbf{P}$ and is given by:

$$J = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m \frac{|\boldsymbol{M}_i|^{1/2}}{\Pr(i)} \exp[-\frac{1}{2}(\boldsymbol{x}_j - \boldsymbol{v}_i)\boldsymbol{M}_i^{-1}(\boldsymbol{x}_j - \boldsymbol{v}_i)^{\mathrm{T}}] + \zeta \sum_{i=1}^{c} \frac{|\boldsymbol{M}_i|^{1/2}}{\Pr(i)} \exp[-\frac{1}{2}(\mathbf{P} - \boldsymbol{v}_i)\boldsymbol{M}_i^{-1}(\mathbf{P} - \boldsymbol{v}_i)^{\mathrm{T}}] \tag{11}$$

where, $\Pr(i)$ is given by (3) and under the constraints:

$$u_{ij} \in [0,1], \qquad i = 1,\dots,c; j = 1,\dots,n$$
$$\sum_{i=1}^{c} u_{ij} = 1, \qquad j = 1,\dots,n \tag{12}$$

The regularization term in equation (11) is zero if $\zeta = 0$ or in the case that all prototypes v_i are equal to $\mathbf{P}$. If the focal point is at a large enough distance from the data, a prototype that is very near the focal point will have values nearing zero in the partition matrix, u_{ij}, this can be deduced by observing equation 4, when the distance $||\boldsymbol{x}_j - \boldsymbol{v}_i||^2$ increases, u_{ij} will tend to zero. These prototypes are considered empty and can be neglected.

The clustering problem can then be formulated as:

$$\mathbf{U}^*, \mathbf{V}^*, \{\boldsymbol{M}_i^*\}_{i=1}^c = \arg \min_{\mathbf{U},\mathbf{V},\{\boldsymbol{M}_i\}_{i=1}^c} J(\boldsymbol{x}; \mathbf{U}, \mathbf{V}, \{\boldsymbol{M}_i\}_{i=1}^c) \quad \text{subject to (12)} \tag{13}$$

Solving the clustering problem can then be accomplished applying Lagrange multipliers λ_j to eq. (11) with constraints (12), which is to minimize:

$$\mathcal{L} = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m \frac{|M_i|^{1/2}}{\Pr(i)} \exp[-\frac{1}{2}(x_j - \boldsymbol{v}_i)M_i^{-1}(x_j - v_i)^{\mathrm{T}}] + \sum_{j=1}^{n} \lambda_j \left(\sum_{i=1}^{c} u_{ij} - 1 \right) + \zeta \sum_{i=1}^{c} \frac{|M_i|^{1/2}}{\Pr(i)} \exp[-\frac{1}{2}(\mathbf{P} - \boldsymbol{v}_i)M_i^{-1}(\mathbf{P} - v_i)^{\mathrm{T}}] \tag{14}$$

The parameter updating expressions are then computed from the respective necessary conditions.

Noticing that the terms that have u_{ij} do not depend on $\mathbf{P}$, the resulting expression will be equal to the non-biased one eq. (4) where the distance metric is (2).

To obtain the estimates of centroids $V = [\mathbf{v}_i] \in \mathbb{R}^{c\times d}$ equation (14) is solved in order to v_i by applying the constraint $\frac{\delta \mathcal{L}}{\delta v_l} = 0$, resulting in:

$$\boldsymbol{v}_i = \frac{\sum_{j=1}^{n} u_{ij}^m \boldsymbol{x}_j + \zeta \mathbf{P}}{\sum_{j=1}^{n} u_{ij}^m + \zeta} \tag{15}$$

Analyzing the previous equation it can be seen that ζ determines the degree of attraction of the prototypes to the focal point, when ζ is large the prototype will tend to $\mathbf{P}$ and if zeta is arbitrarily small the dependency on $\mathbf{P}$ will diminish. Therefore, decreasing the value of ζ results in a finer detailed data observation, which is to say that more clusters are generated, due to less prototypes being attracted to $\mathbf{P}$; while increasing ζ will produce the opposite reaction, the observation will have a smaller number of clusters.

An updating expression for the matrix $\boldsymbol{M}_i$ can be obtained from $\frac{\delta \mathcal{L}}{\delta M^{-1}} = 0$:

$$\boldsymbol{M}_i = \frac{\sum_{j=1}^{n} u_{ij}^m (\boldsymbol{x}_j - \boldsymbol{v}_i)(\boldsymbol{x}_j - \boldsymbol{v}_i)^{\mathrm{T}} + \zeta(\mathbf{P} - \boldsymbol{v}_i)(\mathbf{P} - \boldsymbol{v}_i)^{\mathrm{T}}}{\sum_{j=1}^{n} u_{ij}^m + \zeta} \tag{16}$$

Thus $\boldsymbol{M}_i$ can be viewed as the fuzzy covariance matrix regularized by $\mathbf{P}$. For $\zeta = 0$ it turns into the fuzzy covariance matrix (6).

To avoid numerical errors in the calculation of $\boldsymbol{M}_i$ the following regularizer is employed [11]:

$$\boldsymbol{M}_i^{\mathcal{R}} = \alpha \boldsymbol{I} + \boldsymbol{M}_i \tag{17}$$

4.1. GGFP with focal point in higher dimension

The GGFP, where the focal point is located in a dimension higher than that of the data space, is summarized in Algorithm 1.

Algorithm 1: GGFP with a focal point in a higher dimension of data space

---

Input : Unlabeled multivariate data set: $\boldsymbol{X} \subset \mathbb{R}^d$; Number of clusters: $\mathbf{c}$; Fuzzifier: $\mathbf{m} > 1$; Focal point: $\mathbf{P} \subset \mathbb{R}^w$, $w > d$ Regularization coefficient: $\boldsymbol{\zeta}$ Fuzzy Covariance matrix regularization: $\boldsymbol{\alpha}$

Output: Partition matrix: $\boldsymbol{U} = [u_{ij}]$; Prototypes: $\boldsymbol{V} = [\boldsymbol{v}_i]$; Fuzzy covariance matrix : $\boldsymbol{M}_i$

Initialize the clusters' prototypes ;
Extend $\boldsymbol{X}$ and $\boldsymbol{V}$ to $\mathbb{R}^w$ by adding $(w-d)$ null coordinates per data point;
repeat
  for $i = 1$ **to** c **do**
    Compute $\boldsymbol{M}_i$ using (16);
    Apply regularization $\boldsymbol{M}_i^{\mathcal{R}} \leftarrow \alpha \boldsymbol{I} + \boldsymbol{M}_i$ (17) ;
    for $j = 1$ **to** $|X|$ **do**
      Compute the distance plugging $\boldsymbol{M}_i^{\mathcal{R}}$ in (2);
      Update u_{ij} using (4);
  for $i = 1$ **to** c **do**
    update v_i using (15);
until *a termination criterion was met*;
Project prototypes to the original feature space $\mathbb{R}^d$

4.2. Iterated GGFP

An alternative iterative process to the empirical selection of ζ, proposed in [4] is here applied to the GGFP algorithm and can be seen in algorithm 2. For each value of ζ the resulting clustering partition is evaluated by an internal cluster validation index. In increasing ζ some clusters will be attracted to $\mathbf{P}$, these clusters become neglectable clusters. A cluster is considered neglectable when it does not possess any typical points belonging to the data set. A data point x_j is considered a typical point of a cluster i, iff, the partition matrix for that cluster satisfies the inequality: $u_{ij} > u_{kj}, \quad \forall \quad k \neq i$ [3]. A new partition matrix U is then calculated together with the centroids V and M_i.

5. Results

The observer biased algorithms, FCMFP and GGFP, are compared against each other as well as to their corresponding unbiased versions (FCM and GG). This comparison measures the quality of the resulting partitions using the external validity measure Adjusted Rand Index afterwards, there is also the verification for statistical significant differences between the obtained indexes, this is done using non-parametric statistical tests; the Wilcoxon signed-rank test for comparison of pairs of samples and the Friedman test to compare more than two sets of samples [18].

Algorithm 2: Iterative GGFP

---

input : $\boldsymbol{\Delta\zeta}$, $\boldsymbol{\zeta_{max}}$, $\boldsymbol{C_{max}}$
output: The partition that optimizes the internal validity measure
repeat
  1) Run the GGFP algorithm
  2) Remove negligible clusters using the definition of typicality
  3) Calculate internal validity of clusters
  4) Update $\zeta = \zeta + \Delta\zeta$
until $\zeta = \zeta_{max}$;

Experiments were performed with the planetary gear box of a RCVA-300 vertical axis wind turbine. The experimental setup consists of a blower positioned in front of the wind turbine, a PCB 3-axis accelerometer mounted on top of the gearbox, and a SQI data acquisition system. 10 repetitions per experiment were performed and 155520 data points with 787 features were obtained. The vibrational data was acquired under nominal conditions and for different types of faults in components of the gearbox. A total of 17280 samples were selected for testing. From feature selection the number of 787 features was reduced to 33, using Random Forest feature selection. The data set ready for clustering possesses 17280 data points each with 33 features. there are four classes which are divided as follows: 1) No fault - Healthy state , 2) Ring gear fault, 3) Sun gear fault and 4) Planetary gear.

5.1. Clustering Results

The iterated algorithm 2 is applied in order to obtain a range of reasonable partitions for the data. By tracing a vertical line through both graphs in the location that the validity index is at a local minimum value (for the KL index) or at a local maximum value (for the XB index) the number of clusters that produce that optimal value is obtained.

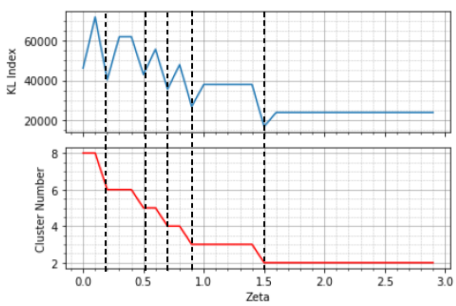


Figure 2: Internal validity index KL and cluster number as function of ζ (Objective is to minimize KL index
)

Observing figures 2 and 3, the KL index reveals as reasonable partitions of data the cluster numbers,

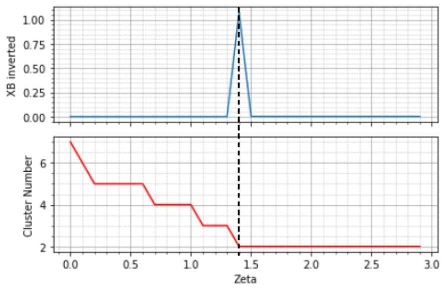


Figure 3: Internal validity index XB inverted and cluster number as function of ζ (Objective is to maximize XB inverted index)

$c = 2, 3, 4, 5$ and 6 while the XB index reveals as reasonable only $c = 2$. The KL index is able to evaluate as reasonable more partitions of the data and also identifies the ground truth partition as reasonable ($c = 4$). Looking at the maximum index value for XB^{-1} and the minimum value for KL it is clear that the partition that optimizes both indexes is the same, $c = 2$; the reason for this can be deduced observing a visual representation of the data, such as figure 5, as the structure in the data seems to be divided in two major parts.

There are ranges of ζ values for which the algorithm gives the same amount of clusters and a similar validation index. Looking more closely at one of those ranges, 30 independent runs with different initial conditions were performed and the performance of the algorithm, in terms of ARI, is tested for the extreme values and two intermediate values of ζ in the range of [0.2, 0.4] that all produce 6 clusters,the results can be seen in figure 4 The Friedman test gives $p_{\text{Friedman}} = 0.6041$

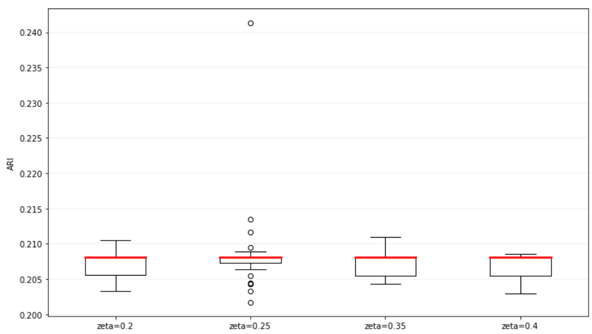


Figure 4: ARI boxplot of GGFP in the range of zeta [0.7, 1.1]

revealing that there are no statistical differences in the data thereby drawing the conclusion that it is possible to change ζ within this range or any other range that produces the same amount of clusters without great change in the ARI values. It is verified here that the GGFP algorithm can give different perspectives of the data without sacrificing performance.

5.2. The healthy vs. faulty case

The data is clustered with c=2 clusters in order to analyse the fault detection case. Figures 5 and 6 show a typical run of the GGFP and FCMFP algorithm for the fault detection case. The different symbols and colors represent the four classes and the red 'x' the prototypes. The ten solid line curves representing a contour of equal membership value (the color mapping from yellow to dark blue colors signifies high to low membership).

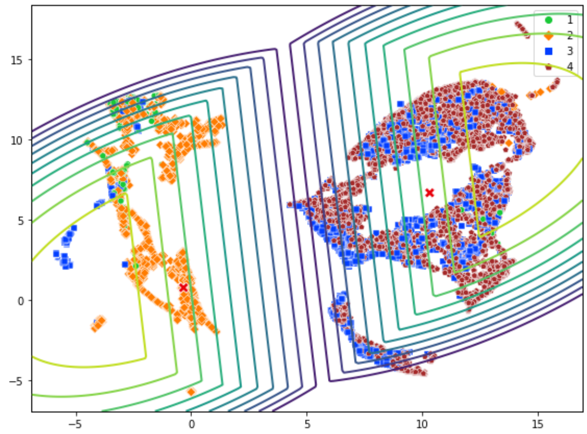


Figure 5: GGFP algorithm UMAP projection of the 33-dimensional feature space into the plane for the fault detection case ($c = 2$) for WT data set

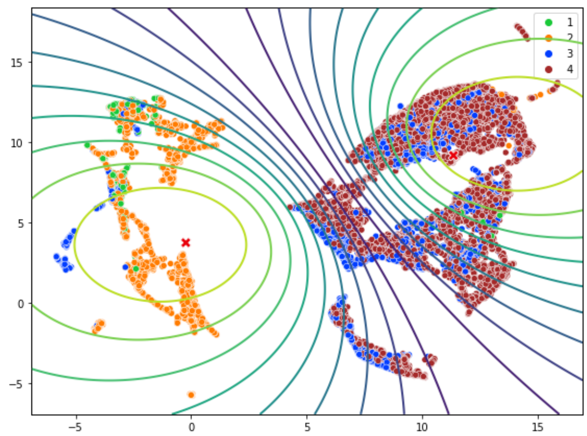


Figure 6: FCMFP algorithm UMAP projection of the 33-dimensional feature space into the plane for the fault detection case ($c = 2$) for WT data set

Analyzing figure 5 it is observable that the majority of the healthy data points (labeled as class 1, represented in green) are clustered together, although there are some mislabeled healthy points.

Most of the faulty points belonging to class 2 and some from class 3 were clustered together with the healthy samples. When comparing the two algorithms visually, it can bee seen the GGFP algorithm managed to cluster most of the healthy samples together and separated more faulty points from the healthy cluster than the FCMFP.

5.3. The multi-fault classification case

Inspecting the cluster validity KL index analysis, in figure 2, partitions with number of clusters $c = 2, 3, 4, 5$ and 6 provide reasonable structural alternatives. Figure 5.3 show the different partitions resorting to UMAP projections in a sequence that appears similar to zooming in on the data. Starting from figure 5, which corresponds to $c = 2$, the cluster where the healthy samples belong gets subdivided into two in figure 7(a), and the trend of the bigger cluster getting subdivided into consecutive sub-clusters is continued in the following figures 7(b) to 7(d).

In order to substantiate the observer metaphor in this data set, the feature space is explored. In figure 8, the focal point is placed in the location where the features obtain their minimum and maximum values for a different number of clusters. The same number of clusters is visualized in each horizontal pair of sub-figures and it is possible to see that different levels of detail, of a certain region, can be obtained depending on the placement of the focal point.

In figure 8, the benefits of being able to control the cluster formation are displayed, for example, if the goal is to analyze the specific region of the data where the healthy samples lie, this could be achieved by placing the focal point in the location where the features obtain their maximum value, allowing the focus to be on a more detailed visualization of the desired area. For example, for the case of $c = 3$, figures 8(a) and 8(b) illustrate this idea: in the right hand side figure there is more detail in the region of the healthy samples when compared with the left-side figure.

5.4. Comparison with the corresponding unbiased algorithm and FCM/FCMFP

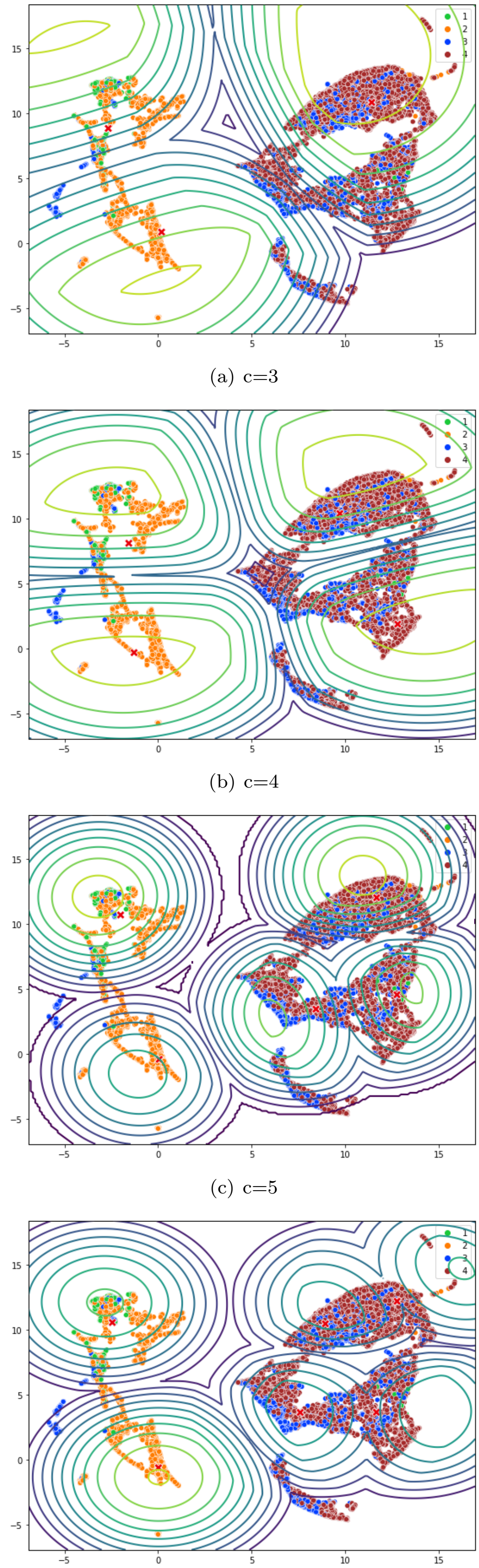
In this section, the algorithms are compared using box-plots of the distributions of Adjusted Rand Index over 30 independent runs of each algorithm, for different numbers of clusters.

The Friedman statistical test revealed a value of $p_{\text{Friedman}} \approx 0$ for all clusters analyzed in figure 9, signifying that there is statistical difference in the data. Performing the Wilcoxon test among the pairs of algorithms revealed that there is no statistical difference in the data produced from the FCM/FCMFP for $c = 2$ and 3 clusters (with $p_{\text{Wilcox}} = 0.0527, p_{\text{Wilcox}} = 0.355$, respectively).

Analyzing figure 9, the GGFP algorithm is able to outperform its unbiased version, as well as the FCM and FCMFP for the reasonable clusters given in figure 2.

It can be observed that the ARI values are higher for the GGFP in comparison to FCM/FCMP for all the analyzed clusters; when compared to its unbiased version, GG, the gap shortens as the number of clusters is increased.

The shrinkage effect of the observer biased framework is again verified and for this data set the benefits of employing the GGFP over the FCMFP or FCM are clear. As for the comparison with its unbiased version, the results show that there is always a slight increase in performance which is more noticeable when the number of



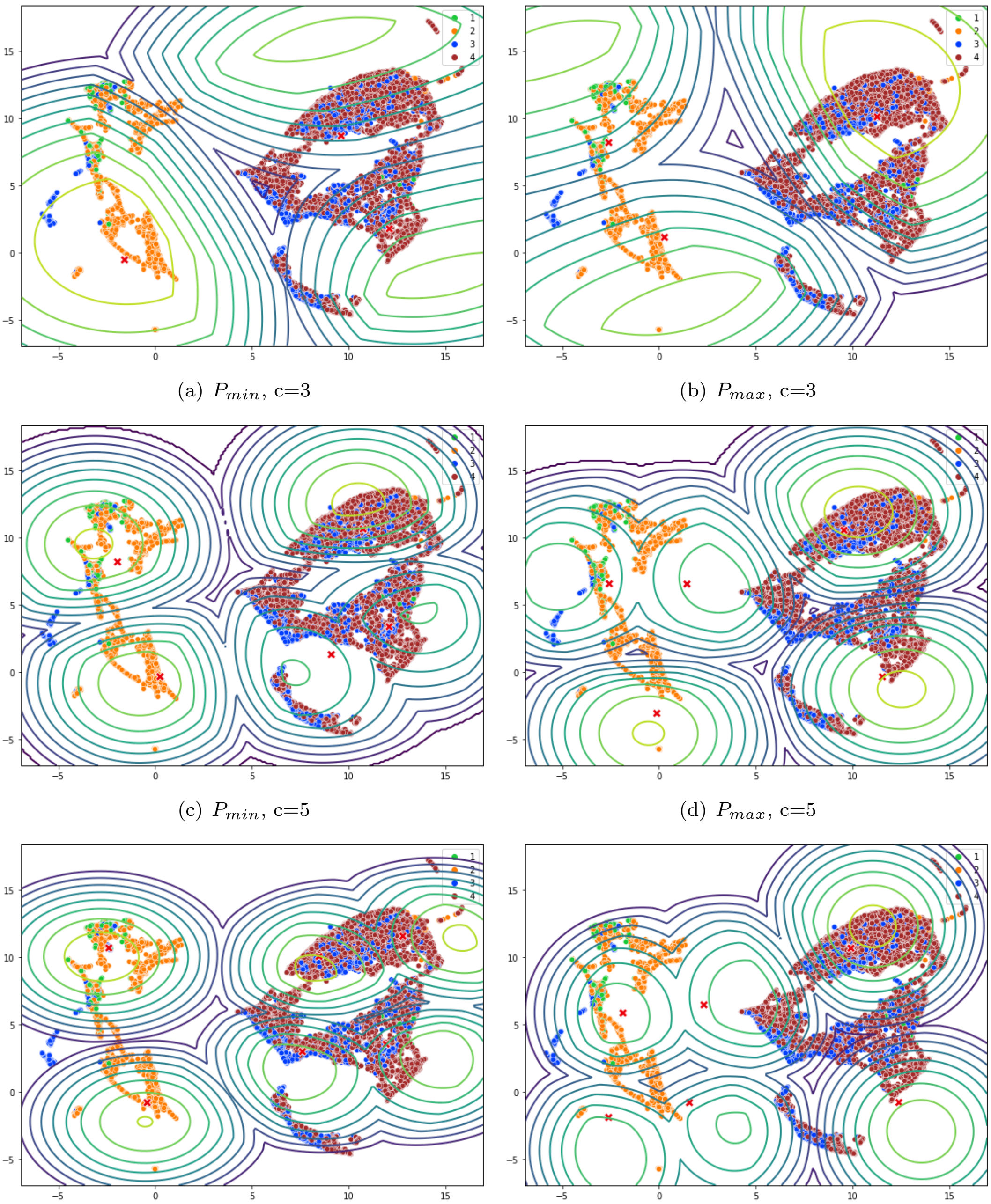
(a) c=3

(b) c=4

(c) c=5

(d) c=6

Figure 7: UMAP projections of the 33-dimensional feature space into the plane for fault classification



(a) P_{min}, c=3

(b) P_{max}, c=3

(c) P_{min}, c=5

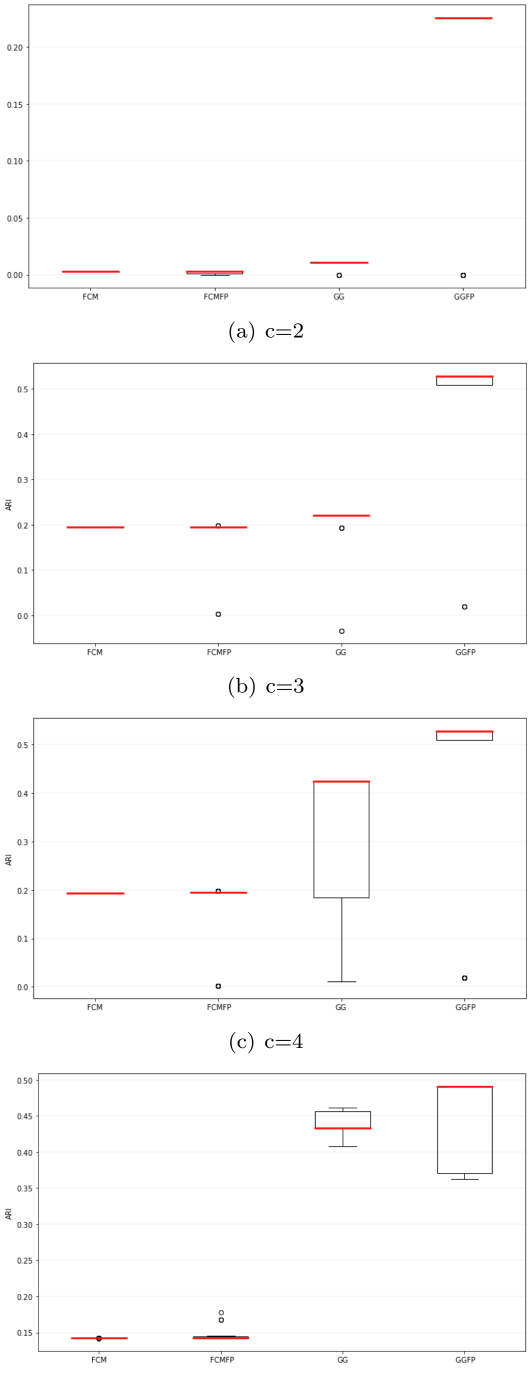
(d) P_{max}, c=5

(e) P_{min}, c=6

(f) P_{max}, c=6

Figure 8: UMAP projections showing more detailed analysis in different regions of the feature space. Left-side sub-figures correspond to P positioned in the region where features attain their minimum values while right side sub-figures correspond to P positioned where features attain their max.

clusters in question are lower. The fault detection case is where the application of the GGFP more notably outperforms the other algorithms.



(a) c=2

(b) c=3

(c) c=4

(d) c=6

Figure 9: Adjusted Rand Index boxplot comparison for different numbers of clusters in the WT data set

6. Conclusions

The GGFP was able to produce statistically significant better results in comparison to the FCM,FCMFP and GG using as a metric the external validity measure, Adjusted Rand Index. The WT gearbox data was of difficult analysis due to the structure of the data which made fault classification a complex task, despite this the GGFP managed to be superior to the other algorithms for all numbers of clusters.

More importantly, the application of the observer biased framework successfully allowed for the exploration of the feature space, making the GGFP a very interesting exploratory analysis tool. The performance increase that the GGFP revealed comes with a price, the fact that the algorithm needs to have a good initialization for the prototypes in order to achieve optimal results makes it a less a robust alternative to the FCMFP as well as the fact that due to its more complex distance metric, the computational time increases considerably. There is a clear trade-off between performance and efficiency and it its up to the user to balance the pros and cons according to needs of the problem at hand.

For future work, the design of a simple and interactive interface could be developed in order to facilitate the exploration of the data space, where the user could be allowed to easily tune parameters such as the location of the focal point or the regularization coefficient. The study of observer-biased clustering with multi-focal points could also be considered and an evaluation on any merits, if any, in applying it over the current version of the observer-biased algorithm.

Acknowledgements

The work was funded in part by national funds through FCT—Foundation for Science and Technology, I.P., through IDMEC, under LAETA, project UIDB/50022/2020, the Prometeo Project of SENESCYT, Ecuador, and the "Multimodal Deep Learning of Deterioration and Fault Prognosis for Wind Turbine Drivetrains" project, funded by Bilateral cooperation between Portugal (FCT: 441.00) and China, through The National Key Research & Development Program, (MOST: 2016YFE0132200). The experimental work relative to bearings was developed at the GIDTEC research group of UPS, Cuenca, Ecuador. The experimental work relative to the wind turbine was developed at the Chongqing Technology and Business University (CTBU).

References

[1] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona. An extensive comparative study of cluster validity indices. *Pattern Recognition*, 46(1):243–256, 2013.

[2] A. Ben Ayed, M. Ben Halima, and A. Alimi. Survey on clustering methods : Towards fuzzy clustering for big data. 2015.

[3] D. Dubois and H. Prade. Fuzzy sets-a convenient fiction for modeling vagueness and possibility. *IEEE Trans. Fuzzy Syst.*, 2:16–21, 1994.

[4] P. Fazendeiro and J. V. de Oliveira. A fuzzy clustering algorithm with a variable focal point. In *2008 IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence)*, pages 1049–1056, 2008.

[5] P. Fazendeiro and J. V. de Oliveira. Observer-biased fuzzy clustering. *IEEE Transactions on Fuzzy Systems*, 23(1):85–97, 2015.

[6] J. Gao and D. Hitchcock. James-stein shrinkage to improve k-means cluster analysis. *Computational Statistics & Data Analysis*, 54:2113–2127, 09 2010.

[7] I. Gath and A. Geva. Unsupervised optimal fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):773–780, 1989.

[8] B. Hahn, M. Durstewitz, and K. Rohrig. Reliability of wind turbines. In *Wind Energy*, pages 329–332. Springer, Berlin, Heidelberg, 2007. doi.org/10.1007/978-3-540-33866-6_62.

[9] L. Hubert and P. Arabie. Comparing partitions. *Journal of classification*, 2(1):193–218, 1985.

[10] Y.-I. Kim, D.-W. Kim, D. Lee, and K. Lee. A cluster validation index for gk cluster analysis based on relative degree of sharing. *Information Sciences*, 168:225–242, 12 2004.

[11] C. Li, L. Ledo, D. Cabrera, R.-V. Sanchez, M. Cerrada, L. Garcia-Hernandez, and J. Valente de Oliveira. A bayesian regularization for the fuzzy covariance matrix of ellipsoidal clusters. In preparation.

[12] Z. Li, R. Jiang, Z. Ma, and Y. Liu. Fault diagnosis of wind turbine gearbox based on kernel fuzzy c-means clustering. *International Conference on Renewable Power Generation*, 2015.

[13] C. L. Liu, X. M. Huang, and X. J. Luo. Roller bearing fault diagnosis based on elmd and fuzzy c-means clustering algorithm. *Applied Mechanics and Materials*, 602:1698–1700, 2014.

[14] S. Lotfan, N. Salehpour, H. Adiban, and A. Mashroutechi. Bearing fault detection using fuzzy c-means and hybrid c-means-subtractive algorithms. *IEEE International Conference on Fuzzy Systems*, 2015.

[15] F. P. G. Márquez, A. M. Tobias, J. M. P. Pérez, and M. Papaelias. Condition monitoring of wind turbines: Techniques and methods. *Renewable Energy*, pages 169–178, 2012.

[16] N. Pal and J. Bezdek. On cluster validity for the fuzzy c-means model. *IEEE Transactions on Fuzzy Systems*, 3(3):370–379, 1995.

[17] K. Pelckmans, J. D. Brabanter, J. Suykens, and B. D. Moor. Convex clustering shrinkage. In *Workshop on Statistics and Optimization of Clustering Workshop (PASCAL)*, pages 2113–2127, London, UK, 2005.

[18] N. R. Smalheiser. Chapter 12 - nonparametric tests. In N. R. Smalheiser, editor, *Data Literacy*, pages 157–167. Academic Press, 2017.

[19] H. Wang, F. Wu, and L. Zhang. Fault diagnosis of rolling bearings based on improved empirical mode decomposition and fuzzy c-means algorithm. *Traitement du Signal*, 38(2):395–400, 2014. https://doi.org/10.18280/ts.380217.

[20] L.-M. Wang and Y.-M. Shao. Crack fault classification for planetary gearbox based on feature selection technique and k-means clustering method. *Chinese Journal of Mechanical Engineering volume*, 31(4), 2018.

[21] E. Wiggelinkhuizen, H. Braam, T. Verbruggen, and L. Rademakers. Condition monitoring for offshore wind farms. 2003.

[22] X. Xie and G. Beni. A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):841–847, 1991.

[23] D. Xu and Y. Tian. A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2, 08 2015.

[24] L. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

[25] Z.Hameed, Y.S.Hong, Y.M.Cho, S.H.Ahn, and C.K.Song. Condition monitoring and fault detection of wind turbines and related algorithms: a review. *Renewable and Sustainable Energy Reviews*, pages 1–39, 2009.

[26] Z.Hameed, Y.S.Hong, Y.M.Cho, S.H.Ahn, and C.K.Song. Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review. *Mechanical Systems and Signal Processing*, 126:662—-685, 2019.

[27] L. Zhang, P. Li, M. Li, S. Zhang, and Z. Zhang. Fault diagnosis of rolling bearing based on itd fuzzy entropy and gg clustering. *Yi Qi Yi Biao Xue Bao/Chinese Journal of Scientific Instrument*, 35:2624–2632, 11 2014.