K- Means Clustering Based on Vibration Signal for Classification Induction Machine Faults

Abla Bouguerne
Department of Electrical Engineering,
University of Constantine 1, Algeria
bouguerneabla@yahoo.fr

Abdesselam Lebaroud
Department of Electrical Engineering,
University of Constantine 3
lebaroud@yahoo.fr

Aziz Boukadoum
Department of Electrical Engineering,
University of Tebessa, Algeria
azizboukadoum@yahoo.fr

***Abstract*— This article present a new method the classification of bearings faults of induction motor. this method using vibration signals of bearing faults normalized by Hilbert transforms for extracted the vectors forms by time-frequency representation dependant class signal ; and recently the essential task that are the raison of our work is the classification of the vectors forms optimized by k- means clustering algorithm. This approach has been tested on vibration data of a test bench composed essentially on a 5.5-kW induction motor.**

Keywords—classification; clustering; time-frequency representation; bearing faults ;Hilbert transform ; k- means algorithm.

I. INTRODUCTION

The fault classification for the three-phase induction machines has been already presented in several publications [1], [2]. In diagnosis of rotating machinery, vibration analysis is widely known to be one of the most effective techniques. This stems from the fact that oscillation is an inherent characteristic of rotating machines and different components of these types of machinery such as shafts, bearings and gears produce vibration energy with different characteristics. Any deterioration in the condition of such components can affect their vibratory attributes and manifest itself in the vibration signature. This allows diagnosis of machine faults by analyzing the vibration signature of the system.

Clustering is traditionally viewed as an unsupervised method for data analysis. However, in some cases information about the problem domain is available in addition to the data instances themselves. In this paper, we demonstrate how the popular k-means clustering algorithm can be profitably modified to make use of this information. Clustering algorithms are generally used in an unsupervised fashion. They are presented with a set of data instances that must be grouped according to some notion of similarity. The algorithm has access only to the set of features describing each object; it is not given any information (e.g., labels) as to where each of the instances should be placed within the partition.

However, in real application domains, it is often the case that the experimenter possesses some background knowledge (about the domain or the data set) that could be useful in clustering the data. Traditional clustering algorithms have no way to take advantage of this information even when it does exist. There are a lot of applications of the K-mean clustering, range from unsupervised learning of neural network, Pattern recognitions, Classification analysis, Artificial intelligent, image processing, machine vision, etc. In principle, you have several objects and each object have several attributes and you want to classify the objects based on the attributes, then you can apply this algorithm.

The k-means method has been shown to be effective in producing good clustering results for many practical applications. However, a direct algorithm of k-means method requires time proportional to the product of number of patterns and number of clusters per iteration. This is computationally very expensive especially for large datasets. we have developed a method using vibration analytical signals of bearing faults normalized by Hilbert transforms for extracted the vectors forms by RTFDCS and is the classification of this vectors forms optimized by k- means clustering algorithm.

II. THE BEARING FAULTS

At the interface between the rotor and the stator, the ball bearing is also have a relatively rapid aging. Typically this type of fault is diagnosed by the spectrum of measurement acoustic or vibration.

Assuming that the gap is not constant during of the bearing faults. this information was also present in the current spectrum, since any eccentricity produces anomalies in the distribution of the field in the gap [9]. the analytical expression of frequencies related to the bearing faults according to its mechanical parameters (number of ball bearings, mechanical speed of the rotor, the contact angle of the balls with the rings ....), and gives the following relationship:

$$f_{bng} = \left| f_s \pm m\, f_v \right| \tag{1}$$

Where: $m = 1,2,3.......$

f_v is a characteristic frequency of vibration.

The expressions for the three types of bearing faults suggests as follows [10]:

Outer race fault :

$$f_{bng} = \left| f_s \pm m\, f_0 \right| \tag{2}$$

Inner race fault :

$$f_{bng} = \left| f_s \pm f_r \pm m\, f_i \right| \tag{3}$$

Ball fault:

$$f_{bng} = \left| f_s \pm f_{frame} \pm m\, f_b \right| \tag{4}$$

The frequencies of bearing faults include rolling the external ring frequency f_0, the interior ring frequency f_i, the balls frequency f_b and the frame frequency [11, 12,13]

$$f_i = \frac{N_b}{2} f_r \left(1 + \frac{Bd}{Pd} \cos\beta \right) \tag{5}$$

$$f_O = \frac{N_b}{2} f_r \left(1 - \frac{Bd}{Pd} \cos\beta \right) \tag{6}$$

$$f_b = \frac{Pd}{Bd} f_r \left[1 - \left(\frac{Bd}{Pd} \cos\beta \right)^2 \right] \tag{7}$$

$$f_{frame} = \frac{f_r}{2} \left(1 - \frac{Bd}{Pd} \cos\beta \right) \tag{8}$$

The five parameters that must be known to calculate the bearing faults frequencies are , the ball diameter Bd , the diameter or average diameter of the frame Pd , the number of balls Nb , the contact angle in radians β , and the rotational frequency f_r .

III. VIBRATORY DATA TREATMENT BY HILBERT TRANSFORM

Vibration Analytics Signal (VAS) method use the principle of the analytical signal obtained by Hilbert transform. The Hilbert transform in the carnal domain corresponds to a phase shift of value $\pi/2$ of all the terms of the Fourier transform. It allows changing the terms cosinus in the terms sinus and the terms sinus in the terms cosinus negative[9].

The Hilbert transform of a signal $y(t)$ can be written as:

$$y(t) \xrightarrow{TH} \tilde{y}(t) = \tilde{y}_{\mathrm{Re}}(t) + j\tilde{y}_{\mathrm{Im}}(t) \tag{9}$$

Where $\tilde{y}_{\mathrm{Im}}(t)$ is the Hilbert transform of the signal $\tilde{y}_{\mathrm{Re}}(t)$, The signal $\tilde{y}(t)$, in turn, is called the analytical signal, we are propos Hilbert function is obtained in the Matlab.

The amplitude modulation $A(t)$ of the carnal signal $y(t)$ is calculated by using the following relationship:

$$A(t) = \sqrt{\tilde{y}_{\mathrm{Re}}(t)^2 + \tilde{y}_{\mathrm{Im}}(t)^2} \tag{10}$$

We can rewrite it in our case about vibration vector $V(t)$:

$$Vt) \xrightarrow{TH} \tilde{V}(t) = \tilde{V}_{\mathrm{Re}}(t) + j\tilde{V}_{\mathrm{Im}}(t) \tag{11}$$

Where $\tilde{V}_{\mathrm{Im}}(t)$ is the Hilbert transform of the vibration signal $\tilde{V}_{\mathrm{Re}}(t)$. The signal $\tilde{V}(t)$, in turn, is called Vibration Analytic signal (VAS)

IV. VECTORS FORMS EXTRACTION BY TFR DEPENDENT CLASS SIGNAL

The discrimination between different classes is made by separating the class i of all remaining classes $\{i+1,\ldots,N\}$ .In this case, The bearing fault kernel is designed to discriminate bearing fault class from the healthy motor class.

The kernels are designed by I training example signals from each class with the equation as follows[1],[2],[3]:

$$J_i(\eta,\tau) = \frac{\left(m_i\,[\eta,\tau] - m_{i-res\tan t}[\eta,\tau]\right)^2}{V_i^2[\eta,\tau] + V_{i-res\tan t}^2[\eta,\tau]} \tag{12}$$

Where:

$m_i[\eta,\tau]$ and $m_{i-res\tan t}[\eta,\tau]$ represent two means of location (η,τ)

$$m_i[\eta,\tau] = \frac{1}{N_i} \sum_{j=1}^{N_i} A_{ij}[\eta,\tau] \tag{13}$$

$$m_{i-remain}[\eta,\tau] = \frac{\sum_{k=i+1}^{4} \sum_{j=1}^{N_k} A_{kj}[\eta,\tau]}{\sum_{k=i+1}^{4} N_k} \tag{14}$$

$I = N_1 \cdot N_2$ is the number of examples per class,

N_2 the number of current examples of same load level and

N_1 the number of load levels,

$V_i^2[\eta,\tau]$ and $V_{i-res\tan t}^2[\eta,\tau]$ represent two variances of location (η,τ):

$$V_i^2[\eta,\tau] = \frac{1}{N_i} \sum_{j=1}^{N_i} \left(A_{ij}[\eta,\tau] - m_i[\eta,\tau] \right)^2 \tag{15}$$

$$V_{i-remain}^2[\eta,\tau] = \frac{\sum_{k=i+1}^{4} \sum_{j=1}^{N_k} \left(A_{kj}[\eta,\tau] - m_{i-res\tan t}[\eta,\tau] \right)^2}{\sum_{k=i+1}^{4} N_K} \tag{16}$$

V. K- MEANS CLUSTERING

Clustering is the process of partitioning or grouping a given set of patterns into disjoint clusters. This is done such that patterns in the same cluster are alike and patterns belonging to two different clusters are different. Clustering has been a widely studied problem in a variety of application domains including neural networks, AI, and statistics[6,7,8].

The k-means method has been shown to be effective in producing good clustering results for many practical applications. However, a direct algorithm of k-means method requires time proportional to the product of number of patterns and number of clusters per iteration. This is computationally very expensive especially for large datasets.

In this section, we briefly describe the direct k-means algorithm [9, 8, 3]. The number of clusters k is assumed to be fixed in k-means clustering. Let the k prototypes $(w_1,\dots,w_k)$ be initialized to one of the n input patterns $(i_1,\dots,i_n)$ .Therefore,

$$w_j = i_l \,,\; j \in (1,\dots,k), l \in (1,\dots,n) \qquad (17)$$

Figure 1 shows a high level description of the direct k-means clustering algorithm. C_j is the j^{th} cluster whose value is a disjoint subset of input patterns. The quality of the clustering is determined by the following error function:

$$E = \sum_{j=1}^{k} \sum_{i_l \in C_j} \left| i_{l.} - w_j \right|^2 \qquad (18)$$

The appropriate choice of k is problem and domain dependent and generally a user tries several values of k. Assuming that there are n patterns, each of dimension d, the computational cost of a direct k-means algorithm per iteration (of the repeat loop) can be decomposed into three parts:

1. The time required for the first for loop in Figure 1 is $O(nkd)$.

2. The time required for calculating the centroïds (second for loop in Figure 1) is $O(nd)$.

3. The time required for calculating the error function is $O(nd)$.

The number of iterations required can vary in a wide range from a few to several thousand depending on the number of patterns, number of clusters, and the input data distribution. Thus, a direct implementation of the k-means method can be computationally very intensive. This is especially true for typical data mining applications with large number of pattern vectors[9,10,11].

---

k-means algorithm

initialize k prototypes $(w_1,\dots,w_k)$ such that

$$w_j = i_l \,,\; j \in (1,\dots,k), l \in (1,\dots,n)$$

Each cluster C_j is associated with prototype w_j

Repeat

for each input vector i_l where $l \in (1,\dots,n)$

do

Assign i_l to the cluster C_{j*} with nearest prototype w_{j*}

for each cluster C_j , where $j \in (1,\dots,k)$, do

Update the prototype w_j to be the centroïd of all samples currently in C_j , so that :

$$w_j = \sum_{i_l \in C_j} i_l / |C_j|$$

Compute the error function :

$$E = \sum_{j=1}^{k} \sum_{i_l \in C_j} \left| i_{l.} - w_j \right|^2$$

Until E does not change significantly or cluster membership no longer changes

---

k-means clustering algorithm

VI. CLASSIFICATION OF VIBRATION BEARING FAULT SIGNAL BY K-MEANS ALGORITHM

A. Data acquisition

The figure 2. represented the dispositive for the signal acquisition: load machine set and signal acquisition. All acquisitions are carried out in continuous over a period of 5 seconds, with a sampling frequency of 20 kHz. Is 100,000 points for each of the measured signals. The Operating modes used to validate the diagnostic procedure are bearing defects were recorded for several load levels 0%, 25%, 50%, 75%, 100%. Each of them is represented by a class in the decision space.

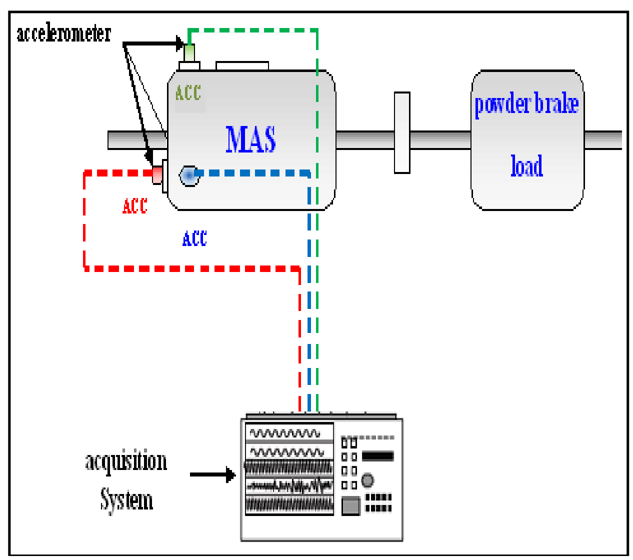

Fig.1. acquisition system of signal

The figure (2) present a vibration data of different bearing faults (axial, radial and vertical) dedicate at induction machine and we are present the periodogram for each data for illustrate the difference between them that shawn in figure (3).

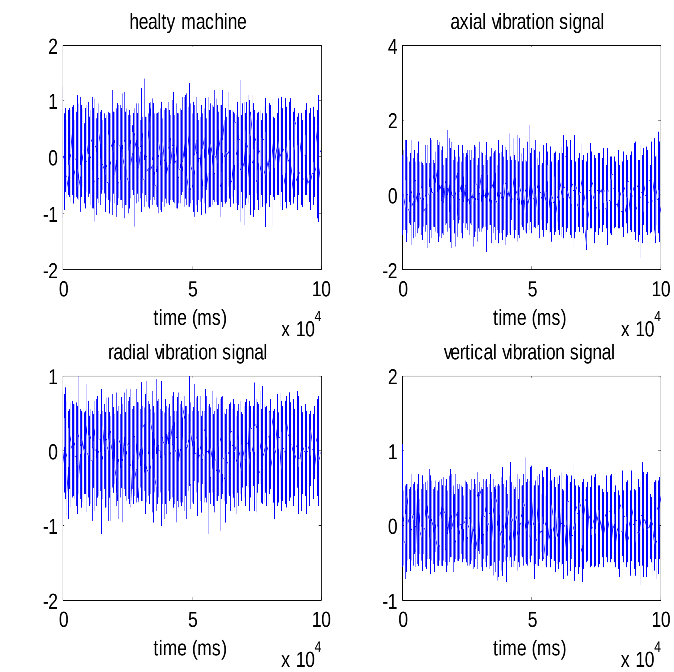

Fig. 2.Vibration data representation

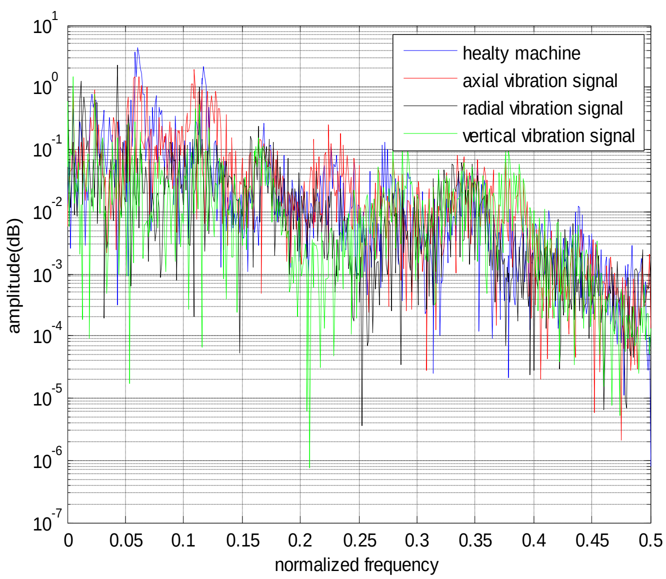

Fig. 3.Periodogram for vibration data

Thus, we also know beforehand that these objects belong to two group's healthy machine and bearing faults (cluster 1 and cluster 2).

The problem now is to determine which Vectors forms optimized belong to cluster 1 and which belong to the other cluster. Each Vectors forms optimized represents one point with two components coordinate.

The figure (4) present the Vectors forms optimized clustering by k- means that is illustrate two classes healthy machine class and bearing fault class.

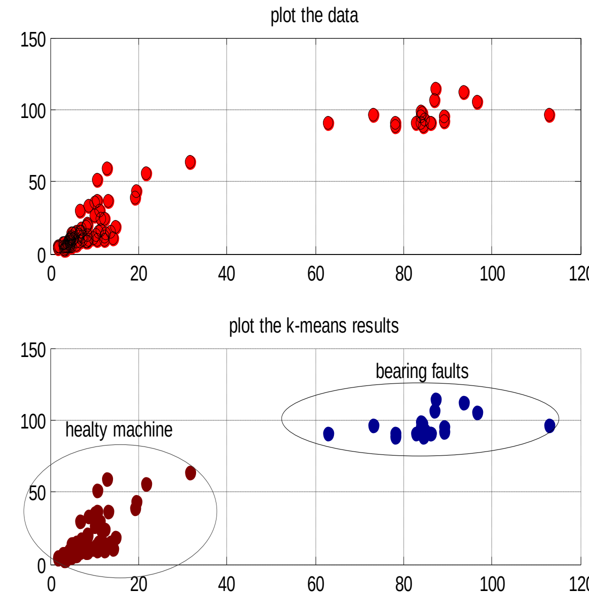

Fig.4. Vectors forms optimized clustering by k- means

Then the K means algorithm will do the three steps below until convergence

Iterate until stable (= no object move group):

1. Determine the centroïds coordinate
2. Determine the distance of each object to the centroïds
3. Group the object based on minimum distance (find the closest centroïd)

The figure (5),(6) shown respectively Data along with the first and final(ten) iteration centroïds obtained by k-means algorithm;

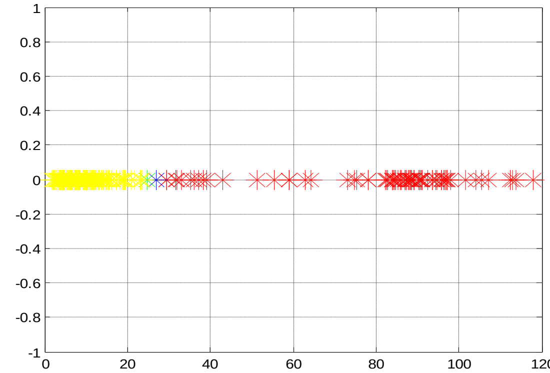

Fig. 5. Vectors forms optimized along with the first centroïds obtained by k-means algorithm

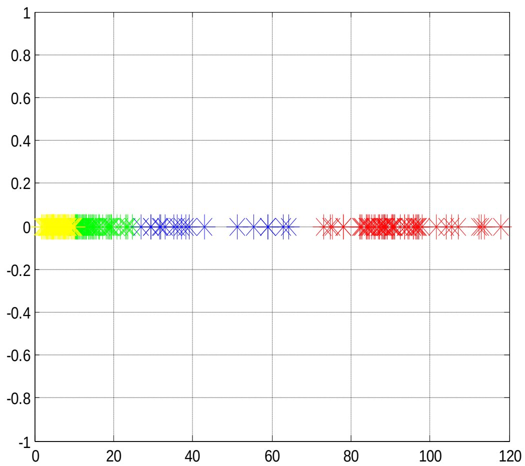

Fig. 6. Vectors forms optimized along with the final centroïds obtained by k-means algorithm

Since we are not sure about the location of the centroid, we need to adjust the centroid location based on the current updated vectors forms optimized. Then we assign all the data to this new centroid. This process is repeated until no vectors forms optimized is moving to another cluster anymore. Mathematically this loop can be proved convergent.

Also present the graph between the sum of the squared difference between the previous membership value and the current membership value (vs.) Iteration number is displayed below in Figure 8. The particular sets of vectors forms optimized (data) are subjected to k-means algorithm Final clusters along with clusters obtained every iteration are displayed in Figure 7. Note that after 6^{th} iteration, clusters are not changed.

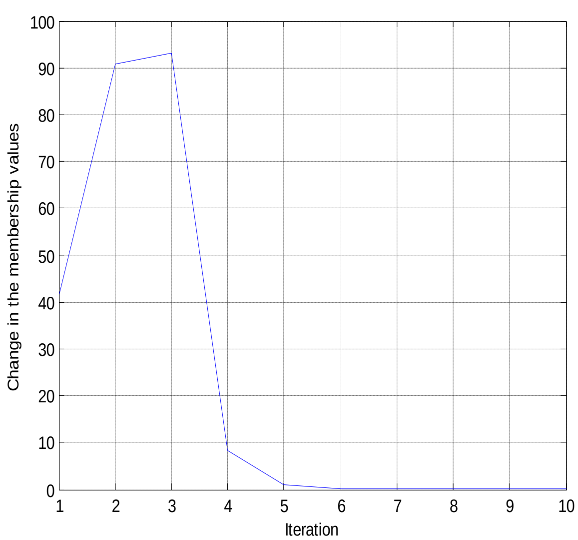

Fig.7. Illustration of change in the membership value every iteration

VII. CONCLUSION

Simply speaking the K-means clustering is an algorithm to classify or to group your objects based on attributes/features into K number of group. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroïd. Thus, the purpose of K-mean clustering is to classify the data.
In this paper, we presented a novel algorithm for performing k-means clustering. Our experimental results demonstrated that our scheme the k-means algorithm can be determine which vectors forms optimized belong to healthy machine class and which vectors forms optimized belong to the other class (bearing faults). Each vectors form optimized represents one point with two components coordinate. The particular sets of vectors forms optimized are subjected to k-means algorithm Final clusters along with clusters obtained every iteration, we note after 6th iteration, clusters are not changed.

VIII. REFERENCE

[1.] O. Ondel, E. Boutleux, and G. Clerc, “Feature selection by evolutionary computing: Application on diagnosis by pattern recognition approach,” in CAINE,S. Dascalu, Ed., 2005, pp. 219–225.

[2.] Lebaroud, Abdesselam, Clerc, Guy "Classification of Induction Machine Faults by Optimal Time-Frequency Representations ", IEEE Transactions on Industrial Electronic, Vol. 57, n°12, pp. 4290-4298, 2008.

[3.] Atlas L., Droppo J., and McLaughlin J., “Optimizing time-frequency distributions via operator theory,” Proc. SPIE, vol. 3162, pp. 161–171, 1997.

[4.] Davy M. and Doncarli C., “Optimal kernels of timefrequency representations for signal classification,” in Proc. IEEE-SP Int. Symp. Time-Freq. Time-Scale Anal., 1998, pp. 581–584.

[5.] Wang M., Rowe G. I., and Mamishev A. V., “Classification of power quality events using optimal time-frequency representations—Part 2: application,” IEEE Trans. Power Delivery, vol. 19, pp. 1496–1503, 2004.

[6.] Costa, Luciano da Fontoura and Cesar, R.M., Shape Analysis and Classification, Theory and Practice, CRC Press, Boca Raton, 2001, pp 577-615.

[7.] K. Alsabti, S. Ranka, and V. Singh. An Efficient K-Means Clustering Algorithm. http://www.cise.ufl.edu/ ranka/,1997.

[8.] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. Introduction to Algorithms. McGraw-Hill Book Company, 1990.

[9.] R. C. Dubes and A. K. Jain. Algorithms for Clustering Data. Prentice Hall, 1988.

[10.] Anderberg, M.R., Cluster Analysis for Applications, Academic Press, New York, 1973, pp. 162-163.

[11.] Schroedl, S., Wagsta_, K., Rogers, S., Langley, P., & Wilson, C. (2001). Mining GPS traces for map re-_nement. (in preparation).

[12.] Fisher, D. (1987). Knowledge acquisition via incremental conceptual clustering. Machine Learning, 2,139{172.