Novelty Detection in Machine Vibration Data Based on Cluster Intraset Distance

Dubravko Miljković
Hrvatska ektroprivreda
Vukovarska 37, 10000 Zagreb
dubravko.miljkovic@hep.hr

Abstract: **Traditional methods for condition monitoring of machinery are based on detecting situations when values of features extracted form measurement data leave predetermined bands consistent with normal machine operation. Design of such systems requires considerable amounts of measurement data describing machinery failure modes that are generally very difficult to obtain due to large number of failure examples necessary. Novelty detection is the identification of new or unknown data or signal that a machine learning system is not aware of during training. Assumption is that in case of impeding failure, new previously unseen measurement data will appear. In this paper method of novelty detection in machine vibration data is described based on clustering of features extracted from measurement data. During training system discover main operational regimes of machine and assign to them clusters of feature data. Later, during machine exploitation, by comparing intraset distances within cluster members with closest distance of new example to cluster centers, system is able to detect abnormal new measurement data that were not known at the time of training the model. Method is presented on car engine vibration data.**

Index Terms: **novelty detection, vibration analysis, clustering**

I. INTRODUCTION

Machine operation involves the generation of forces that produce vibrations. During operation, all machines are subjected to fatigue, wear, deformation, and foundation settlement. Vibration signature of an operating machine provides a lot of information about the inner working of the machine. When a vibration problem exists, a detailed analysis of vibration can identify the specific cause. Vibration analysis is used today as a predictive maintenance tool in a wide variety of industrial areas, especially for rotating and reciprocating machines. For successful detection of machine problems by vibration analysis it is necessary to know vibration signatures of a machine in good running condition. When confronted with a large pool of machines that were in use for a long period it is possible and convenient to collect vibration signatures sampled at successive intervals and form a database for future vibration analysis. However, when machine is of a new kind, put recently into exploitation, very reliable, or pool of machines is very small, there are simply not enough examples of faults, as they either don't exist yet, or are very rare, and hence there are not enough vibration signature data for comparisons. In such situations novelty detection in vibration data may be viable solution for machine vibration monitoring.

II. NOVELTY DETECTION

Novelty detection is the identification of new or unknown data or signal that a machine learning system is not aware of during training. Instead of training the system to recognize the faults, the system learns a model of the normal environment that does not have any problems and the novelty filter detects deviations from this model. A vibration signature constructed from many hours of data from normal machine operation may be used as a model of normality for that machine. New values observed during machine operation are compared to the model. Decision whether new values are normal or abnormal is based on decision boundary or novelty threshold. The objective of novelty detection is the generation of reliable and robust alerts if the condition of the system being monitored is deemed to have deteriorated, [1]. Novelty detection is generally performed using statistical methods and artificial neural networks (eg. Kohonen self-organizing maps - SOM, Radial Basis Functions - RBF). Overview of novelty detection methods can be find in [1, 2, 3].

III. CLUSTERING

A. BASIC CONCEPTS

Clustering is the partitioning of a data set into subsets (clusters), so that the data in each subset share some common trait. In clustering process is necessary to select a distance measure, which will determine how the similarity of two elements is calculated. Similarity is related to distance in the sense that the greater the similarity between two data points, the lesser the distance is between them.

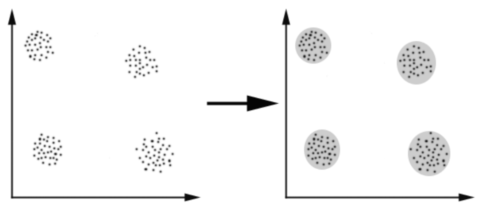


Fig. 1. Illustration of clustering

Clustering based approaches are aimed at partitioning data into a number of clusters, where each data point can be assigned a degree of membership to each of the clusters. If the degree of membership is thresholded to suggest if a data point belongs or not to a cluster, novelty can be detected when a sample belongs to none of the available classes, [4].

Two important concepts in clustering are:
- Interset distance: distance between members of same cluster.
- Intraset distance: distance between members of different clusters.

B. K-MEANS CLUSTERING

K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. This algorithm aims at minimizing an objective function, in this case a squared error function. The objective function is

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \tag{1}$$

where

$$\left\| x_i^{(j)} - c_j \right\|^2 \tag{2}$$

is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, is an indicator of the distance of the *n* data points from their respective cluster centres (centroids). Domain knowledge must be used to guide the formulation of a suitable distance measure for each particular application.

The basic k-means algorithm consists of the following steps:

1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

Results of clustering process are dependent on choice of initial cluster centers. The main advantages of this algorithm are its simplicity and speed which allows it to run on large datasets. Its disadvantage is that it does not yield the same result with each run, since the resulting clusters depend on the initial random assignments. It minimizes intra-cluster variance, but does not ensure that the result has a global minimum of variance.

IV. NOVELTY DETECTION USING CLUSTERING

Clustering is suitable for creating dataset clusters that correspond to numerous regimes of normal operation. Novelty detection method works in two modes: learning and operating mode. During learning mode description of normality is acquired by clustering process on presented data that correspond to normal machinery operation. In operating mode new signal is verified for normality by determining similarity with formed cluster representatives.

A. LEARNING MODE

Clustering process can be performed using k-means clustering. Learning can be unsupervised if training data is not labeled or supervised if there are labels attached to available training data learning. In later case clustering algorithm can be used to find new cluster representatives different from initial choices, but also to move some data from one cluster to another as these data may better be members of other cluster. Such case can result from engine regime transitions captured among data for particular regime. Training data associated with normal machine operation includes data from several regimes of machine operation. Each engine regime corresponds to a cluster center.

B. OPERATING (MONITORING) MODE

New samples (ie. test feature vectors) are compared with cluster representatives (cluster centers). Closest mach is find and distance compared to average intraset distance from cluster center to other members of same cluster. If distance of new sample surpass novelty threshold, sample is considered as a novelty and warning alarm is given.

C. NOVELTY THRESHOLD

There exist various ideas how to set value for novelty threshold. If distance is smaller or equal to threshold test feature vector is considered as normal:

$$d_T \le T \tag{3}$$

If larger than a threshold test feature vector is considered novelty (abnormal situation):

$$d_T > T \tag{4}$$

Some common ways for calculation of the threshold are:

1. maximum training distance for particular cluster

$$T = \max_i d(x_i, c) \tag{5}$$

2. calculate mean μ and standard deviation σ of training distances for particular cluster

$$T = \mu + n\sigma \tag{6}$$

Assumption is that if distribution of training distances is normal, distance of few σ from cluster centroid enclose almost all possible members for that cluster (in case of normal distribution $\pm 3\sigma$ account for 99.7% of all data). Training distance is here considered as a distance between cluster centroid and cluster member.

3. In addition new sample can be considered a novelty if any feature vector element surpass minimal or maximal value of that element present in cluster training data. Feature vector is considered a novelty if there exist vector element outside minimal and maximal boundaries for that element within cluster:

$$\exists x_i \quad x_i > x_{\max} \tag{7}$$

or

$$\exists x_i \; x_i < x_{\min} \tag{8}$$

x_{min} is minimal value of element x_i within particular cluster, and x_{max} is maximal value of element x_i within particular cluster

4. Novelty detection via Cluster Win Frequency

In some cases all test feature vectors classified to some cluster can be recognized as novel. This may be due to low training win frequency for that cluster or specific knowledge of the problem domain. Any new feature vector that maps to that cluster is considered novel.

V. NOVELTY DETECTION IN MACHINE VIBRATION DATA

In order to store, display, and modify vibration signals on a digital computer, the signals must be digitized. This is done through two processes known as sampling and quantization, process known as analog-to-digital (A/D) conversion.

A. DATA PREPROCESSING

Data preprocessing includes include following activities:

1. Adjusting input level - amplitude normalization
2. Temporal processing - downsampling in case of very high sampling rate
3. Frequency processing - low pass filtering that cut signal frequency components above half of sampling rate.
4. Windowing is generally not performed for time-domain analysis (except special case of rectangular window), but is common for frequency analysis

B. FEATURE EXTRACTION

Feature extraction is the representation of signals using a smaller set of quantities, termed features [3]. In analysis of machine vibration data time domain features have been used: RMS, crest factor and kurtosis. This is just one educated choice, best set of features is application dependent. For particular class of faults features in frequency domain may be more suitable.

1. RMS - Root Mean Square value

The RMS value of a vibration signal is an important measure of its amplitude.

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x_n^2} \tag{9}$$

The crest factor and kurtosis function are used to describe the shape characteristics of the signal and change together with vibrations produced by machinery.

2. Crest factor C_f is defined by*:*

$$C_f = \frac{P_L}{RMS} \tag{10}$$

P_l is the signal peak level.

The crest factor calculation tells us how much impacting is occurring in a waveform.

3. Kurtosis

The fourth order moment, kurtosis, is defined as:

$$k = \frac{\sum_{n=1}^{N}\left[y(n) - \mu\right]^4}{N\left(\sigma^2\right)^2} \tag{11}$$

The kurtosis is sensitive to impulsiveness or "spikeness" of the data.

C. CONSTRUCTING A FEATURE VECTOR

Feature vector is constructed for each successive time interval (eg. every 50 ms).

$$x_n = \left[x_n^1 \ldots x_n^D\right] \tag{12}$$

After some time we get set of N feature vectors

$$\{x_1, \ldots, x_N\} \tag{13}$$

for time intervals *1 ... N* comprising one dataset.

D. NORMALIZATION

Different features in a feature vector generally don't have same dynamic range. It is desirable to normalize feature vectors in such way that individual elements may be compared. Among many methods for normalizing a feature vectors, one appropriate choice may be component-wise normalization described in following equation:

$$x_n' = \left[\frac{x_n^1 - \mu_1}{\sigma_1}\frac{x_n^2 - \mu_2}{\sigma_2}\cdots\frac{x_n^D - \mu_D}{\sigma_D}\right] \tag{14}$$

where

$$\mu_d = \frac{1}{N}\sum_{i=1}^{N} x_i^d \tag{15}$$

is the mean, and

$$\sigma_d = \sqrt{\frac{1}{N-1}\left(x_i^d - \mu_d\right)^2} \tag{16}$$

standard deviation of element x_i^d across the whole dataset. If all elements of a features vector have same dynamic range and are of equal importance in classification decision Euclidian distance may be convenient choice as similarity measure between elements during clustering and classification process.

$$d_{i,j} = \sum_{k=1}^{D} \left(x_i^k - x_j^k\right)^2 \qquad (17)$$

E. CLUSTERING OF TRAINING DATA

After feature normalization clustering process is performed using k-means algorithm. Number of cluster centers corresponds to number of engine regimes for which training data is available. Initial cluster centers (centroids) are determined for each training data subset (engine regime) by determining sample mean of labeled training set.

$$\mu_d = \frac{1}{N} \sum_{i=1}^{N} x_i^d \qquad (18)$$

Feature vector x_i closest to mean feature vector is considered starting centroid, c.

$$d_{\substack{i \\ \min}} = \sum_{k=1}^{D} \left(x_i^k - \mu_k\right)^2 \qquad (19)$$

$$c = x_i \qquad (20)$$

For some regimes there are more training data subsets, for each subset there is one centroid. After choice of initial cluster centroids is finished, k-means algorithm is performed on all training data and new cluster centroid and cluster members are determined.

F. CLASSIFICATION OF NEW FEATURE VECTOR

For each new feature vector, closest cluster centroid in terms of similarity by Euclidian distance is determined. In such way new vector is classified to cluster that corresponds to one engine regime. Than, similarity measure is compared to novelty threshold. If Euclidian distance is greater of novelty threshold, than feature vector is considered as novelty.

VI. EXPERIMENT SETUP

Signals have been downloaded from internet [5], but were initially collected from car engine. Details of car end engine are following:

- Car: 1979 Chrysler Sigma GL, Built in Adelaide, South Australia
- Engine: Astron 2000 engine - 2 litre, four cylinder with balancing shafts
- Idle Speed 680 rpm

A. DESCRIPTION OF VIBRATION SIGNALS

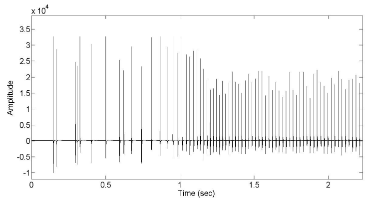
Seven samples of the vibration of the car engine were taken by affixing an accelerometer to the air cleaner. The signal from the accelerometer was captured using a Soundcard, and recorded as a .wav file. The samples were recorded as .wav files with 16 bit A/D conversion and a sample rate of 22 kHz. Data collected represent four engine regimes with seven signal examples:

1. Car Engine (Starting): Signal 1
2. Car Engine (Idling) - two signal examples (cold and worm engine): Signals 2 and 3
3. Car Engine (Revving under Load) - three examples (different rotation speeds): Signals 4, 5 and 6
4. Car Engine (Stopping): Signal 7

Later, during learning process these seven signals will produce seven training datasets with data from normal engine operation.

Collected vibration signals are presented in time and frequency domain in figures 2 to 5. The frequency spectra were obtained by applying a Hanning window to the time signals and then using an FFT with 2048 points to transform them to the frequency domain. Although some of the time signals for an object appear quite different, the spectra obtained using the FFT shows the basic information content of the signal is the same.

a)



b)

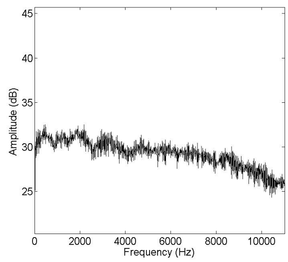
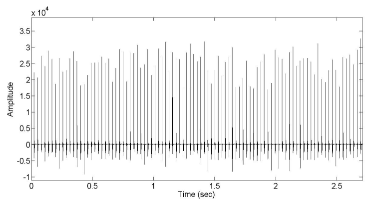


Fig. 2. Starting: a) waveform b) spectar

a)



b)

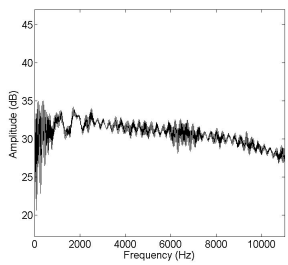


Fig. 3. Idle 1: a) waveform b) spectar

a)

b)

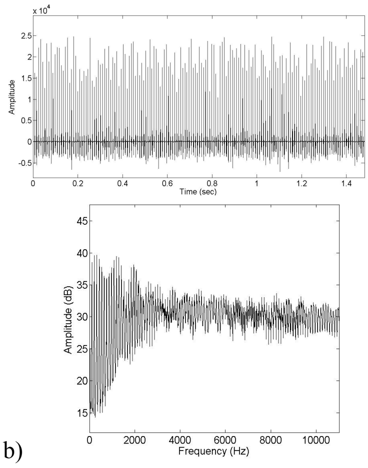


Fig. 4. Rev 1: a) waveform b) spectar

a)

b)

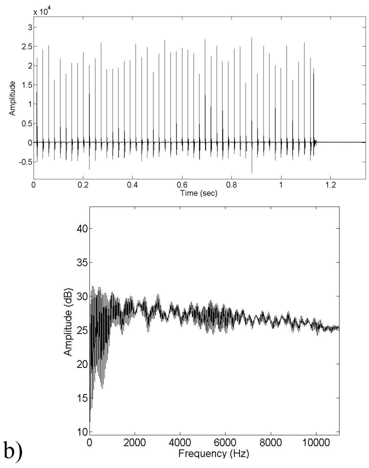


Fig. 5. Stopping: a) waveform b) spectar

B. LEARNING MODE

Feature vectors have been extracted from recorded signal in intervals of 50 ms. After normalization and determination of initial cluster centers clustering process has been performed. Downloaded engine vibration data were used as training data. Feature vector elements were of different dynamic range and data normalization was performed according to (14). Mean values from each training dataset have been used as initial cluster centers. K-means clustering method was performed on presented training datasets and new cluster centers were determined. Cluster centers remained stable after 10 iterations. For each cluster center distances to all other members of same cluster have been determined and following statistics calculated, Table I.

TABLE I DISTANCES CLUSTER CENTER TO MEMBERS

Cluster	Distances from cluster center to other cluster members		
	μ	σ	max
1	0,562389	0,31446	1,036689
2	0,507966	0,251336	0,863775
3	0,557343	0,252240	1,171913
4	0,432615	0,254267	1,044999
5	0,385927	0,167391	0,704329
6	-	-	-
7	0,624355	0,232188	1,137706

In cluster six there was just one element, used as cluster center.

C. MACHINE OPERATION (MONITORING)

The nearest neighbor technique represents a very practical approach for direct classification. For each new feature vector, classification is performed using 1-NN rule. Distance to the closest member of that set is compared against threshold. In experiment two types of thresholds were used (yielding same result on used test datasets). Threshold can be mean intraset distance from cluster center augmented by three standard deviations:

$$T = \mu + 3\sigma \qquad (21)$$

or in simpler case, not taking account of cluster statistics, largest intraset distance between cluster center and all members of that set augmented by 20%:

$$T = 1.2 d_{\max} \qquad (22)$$

Maximal distances for each cluster center are presented in Table II.

TABLE II THRESHOLDS FOR VARIOUS CLUSTERS

Cluster	μ+3σ	1.2 d_{MAX}
1	1,505769	1,2440268
2	1,261947	1,03653
3	1,314063	1,4062956
4	1,195416	1,2539988
5	0,8881	0,8451948
6	-	-
7	1,320919	1,3652472

Because cluster center is the only element of cluster 6, the threshold distance was not determined.

Testing process was performed in two ways:

1. Testing with previously unseen normal data (part of engine data)

In all training datasets last feature vector was removed from training process, and reserved for testing purpose. Such test feature vector represents known normal situation. There were seven training datasets available from which was taken total of seven test features vectors.

2. Testing with previously unseen abnormal data (synthetic sample)

Unfortunately, there was no easily available car engine vibration data with fault example. To test the method anyway, original signal corresponding to normal engine operation (revving under load) was distorted by one of many functions of CoolEdit program (dynamic range processing), to some extent equivalent to engine operation where vibration impulses are much more pronounced. Distorted waveform and corresponding spectrum are shown in Fig. 6.

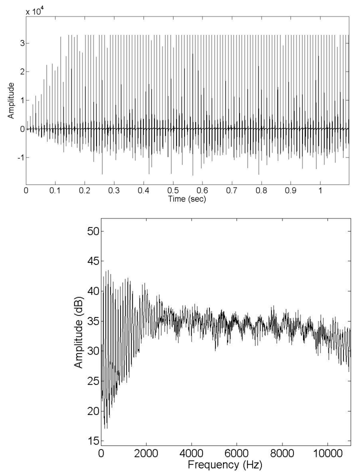


Fig. 6. Synthetic fault: a) waveform b) spectar

D. *RESULTS OF TESTS*

Results of tests are presented in Table III for unseen normal training data and in Table IV for synthetic abnormal data. For available testing data same results were achieved for both variants of threshold.

TABLE III CLASSIFICATION OF NORMAL DATA

Feature vector	Classified as dataset	Distance	Normal /Abnormal
1	7	0,439109	N
2	4	1,180472	N
3	2	0,397505	N
4	5	0,148827	N
5	5	0,202249	N
6	4	0,215940	N
7	6	0,017204	A (?)

TABLE IV CLASSIFICATION OF ABNORMAL DATA

Feature vector	Classified as dataset	Distance	Normal/Abnormal
1	5	0,403693	N
2	4	0,963157	N
3	4	2,602669	A
4	4	2,829521	A
5	4	3,270817	A
6	4	3,856194	A
7	4	4,710083	A
8	4	5,012846	A
9	4	2,744308	A
10	4	4,845574	A
11	4	3,742057	A
12	4	3,000267	A
13	4	3,949844	A
14	4	3,347194	A
15	4	3,834809	A
16	4	4,928098	A
17	4	2,900509	A
18	4	4,779285	A
19	4	1,992792	A
20	4	5,126909	A

Results of tests are summarized in Table V.

TABLE V SUMMARY OF TEST RESULTS

Testing Data	Classified as	
	Normal	Abnormal
Training(Unseen)		
7 test vectors	6 or 85,7(100*) %	1 of 7 or 14,3(0*) %
Synthetic Fault		
19 test vectors	2 of 19 or 12%	17 of 19 or 82%

*If cluster with just one element is discarded from consideration as an exception, all test vectors are correctly classified

Feature vectors from previously unseen normal data have been correctly classified as normal in six of seven cases. One feature vector that has been classified as abnormal came from engine start sequence that is not so usual in engine work and closest cluster representative was sole member of that cluster (other clusters had 9-29 feature vectors). Synthetic produced abnormal data have been correctly classified as abnormal in 17 of 19 cases of presented feature vectors. Reliable statistics would require much large dataset of unseen normal data, however original datasets have been small and no more feature vectors could be separated for testing purpose as they were also necessary in a training process.

VII. CONCLUSION

Presented is simple, yet powerful and intuitively acceptable method for novelty detection. Described experiment serves as a practical illustration of this concept. For real-world application it would be necessary to work with much larger training datasets for improved accuracy. Because comparison of test feature vectors is performed only with cluster representatives (cluster centers) method is not computationally intensive during monitoring (operational) phase.

REFERENCES

1. M. Markou and S. Singh, Novelty Detection: A Review Part I: statistical approaches, Signal Processing, Vol. 83, Issue 12, December 2003
2. M. Markou and S. Singh, Novelty Detection: A Review Part II: neural network approaches, Signal Processing, Vol. 83, Issue 12, December 2003
3. L. Tarassenko, D.A. Clifton, P.R. Bannister, S. King, D. King, Novelty Detection, Encyclopedia of Structural Health Monitoring. Wiley, 2008. In press.
4. P. Duke, Cluster-based Methods for Novelty Detection, Predictive Dynamix Inc.
5. R. Roberts and G. Roberts, The Vibration Web Page, Analyisis od Example Vibration Data www.cage.curtin.edu.au/mechanical/info/vibrations/welcome.htm