Application of Classification Analysis for Monitoring Knock Intensity in Gas-Fueled Marine Engines

by

Oleksiy BONDARENKO*, FUKUDA Tetsugo*

Abstract

Knocking is an abnormal combustion phenomenon that occurs in gas-fueled engines. It is a problem that needs to be addressed due to its impact on the efficiency and safety of engine operation. Knock combustion is caused by the nonstationary and stochastic ignition of the air-fuel mixture outside of the flame front. The frequency band of the resulting resonant pressure waves differs from those of normal combustion. This property is highly suitable for the application of clustering and classification methods from machine learning (ML). This study elaborates on the application of ML techniques, the support vector machine (SVM) and the self-organising map (SOM), for monitoring knock onset. The application of both requires specific preprocessing and post-processing steps. Feature engineering, providing for the selecting of attributes that characterise the state of combustion in the engine is especially of great importance. A previously developed algorithm for feature extraction was used to justify the application of the ML algorithms. Furthermore, the preprocessing step for the SVM algorithm requires supervised learning, so fuzzy c-means (FCM) clustering was used to assign labels to clusters. Similarly, the post-processing step provided the similarity measure between the model vectors of the SOM and a new features sample, so the FCM algorithm was added on top of the SOM. The algorithms for classifying knock intensity were applied to an engine with controlled knock conditions, and the results were comparable to those obtained by a classical approach, with the added benefit of continuous monitoring. Furthermore, the efficacy of ML techniques was studied with respect to identifying knock combustion on the basis of the measured engine acoustic emission.

＊ 環境・動力系
原稿受付　令和３年４月２３日
審 査 日　令和３年６月１５日

Contents

1. Introduction

1.1. Background

A series of ambitious strategies toward greener shipping has been recently initiated in a new Resolution MEPC.304 (72) adopted by the International Maritime Organisation (IMO). Along with the continuing intense pressure to limit nitrogen oxides (NOx), sulphur oxides (SOx), and particular matter (PM) from ship exhaust gasses spur the ship-owners to consider liquefied natural gas (LNG) as an alternative fuel of choice.

In 2017, the in-service and on-order fleet of LNG-powered seagoing vessels (excluding LNG-carriers) had reached the 200 ships milestone showing the year-to-year jump of over 20% [1]. Such a growing trend specifies new opportunities as well as technical challenges, especially regarding abnormal combustion control system. Unlike the traditional Diesel engines, both the gas-fueled and dual-fuel engines are prone to the occurrence of abnormal combustion, called 'knock'. The knocking phenomenon is mainly caused by the spontaneous ignition of the air-fuel mixture in the unburned zone (end gas) before the flame front, initiated by a spark plug, arrives. This produces a shock wave that generates rapid increase and oscillation of in-cylinder pressure. It results in a decrease in engine power, increased fuel consumption and wear of the engine, and in the worst case, it may cause damage to the engine components. Simultaneously, the highest efficiency of the gas-fueled engine is achieved on the edge of the self-ignition or knock margin. In order to achieve an ultimate engine performance preserving the safe operation, an efficient control system is required. In turn, the need for a continuous assessment of the state of knock intensity is an essential condition for the control to be performed.

1.2. Knock identification approaches

For many decades, knock has been recognised as a significant problem receiving a vast amount of attention to understand the phenomenon better and aimed at proper identification and control. The best and most accurate information on knock is retrieved from the measured in-cylinder pressure as it is accompanied by the shock waves from the self-ignition spots within the cylinder charge [2, 3, 4]. On the other hand, the shock wave produced by the detonation induces vibration of the engine structure that can be detected by using vibration sensors [5, 6]. Similarly, acoustic emission (AE) from the engine also contains components related to the detonation and can be used to identify the occurrence of knocking [7, 8]. In order to detect knock, the information from the sensors is subjected to a time and frequency-based analysis as well as a joint time-frequency analysis (Short-Time Fourier Transform, Wavelet, etc.). In turn, the effective identification of the onset and tendency to knock by these methods is provided by ensuring a high signal-to-noise ratio and adopting an appropriate threshold. However, the sustainable and continuous detection of knock intensity is limited due to the apparent stochastic behaviour and nonstationary spectral content of the knock phenomenon.

Because of the high nonlinearity and nonstationarity of the knock identification problem, machine learning (ML) approaches have been found promising. The ML is used to develop an information model of knocking, which is essentially a black-box model. This implies that the computational intelligence methods are used to formalise the relationship between the features in the available engine measurements and the condition of knock onset. Concerning the problem of knock detection, a subset of clustering methods is of primary interest.

The Support Vector Machine (SVM) is the method from the statistical learning theory that intends to solve classification issues and is found to provide fairly well general performance in classification and detection of knocking [9]. However, the SVM is the supervised learning algorithm, and this implies that the training has to be performed on the labelled data set where each class label corresponds to some feature obtained from either in-cylinder pressure or vibration signals.

Another widely used tool for unsupervised classification is a kind of Neural Network (NN) called Self-Organising Map (SOM). The SOM network is mostly used to represent the structure and dependencies in the multivariable data. Thus, the quantisation of the time-frequency properties of the data obtained from the engine sensors can be performed using the model vectors of the SOM map neurons, and then the map is used to account for knock or no-knock events at every cycle [10]. The SOM, however, gives no probability measures [11] to determine if the sample belongs to the information model determined by map neurons. Thus, the SOM classification alone cannot explain the intensity of the abnormal condition.

1.3. A hybrid approach for knock intensity monitoring

In recent time, the trend in artificial intelligence systems is to develop a hybrid system combining the supervised and unsupervised learning approaches [12] whereby constructing a superior performance system. Furthermore, proper data preprocessing and feature engineering mainly drive the model performance, and these are crucial for subsequent fidelity improvement. In the preceding study [17], a sophisticated signal processing algorithm was developed, providing for the discrimination of every combustion cycle as being non-knocking or knocking. Furthermore, the algorithm is highly suitable for processing indirect combustion-related data such as acoustic emission (AE) where the correlation between raw data and knock intensity is not apparent. Thus, the knock intensity monitoring problem can be seen as the problem of identifying to which of a set of categories (knocking or non-knocking) an observation (features derived from every combustion cycle) belongs to. The latter implies that the clustering and classification algorithms of ML are highly suitable for the task of knock intensity monitoring. Therefore, it is proposed to combine the feature extraction algorithm with the combination of ML techniques. Notably, the features extracted from the signal are then processed with a Fuzzy c-Means (FCM) clustering algorithm to assign labels to clusters of raw features. Further, the two-class SVM classifier is trained and used to explain the intensity of the knock. In much the same way, the SOM is trained over a set of raw features to map the topology and then the FCM model built on top to provide the probability measure and thus is used to classify the knock intensity.

2. Data and Method for Feature Extraction

2.1. Data collection

A vast database was collected during past researches on the knocking characteristics of the gas-fueled engine. Table 1 summarises the specification of the used lean-burn gas engine. During the experiments, Butane base LPG (Liquefied Petroleum Gas) was mixed with city gas (90% Methane) in order to change supplied fuel gas composition and provoke knocking condition. The engine parameters, such as ignition timing and charge air temperature, were also concurrently subjected for tuning exciting the knock of varying severity. The in-cylinder pressure data was acquired from a pressure sensor installed in one cylinder and synchronised with the crankshaft position with a resolution of 0.5 degrees. Besides, a microphone was set in close proximity to the cylinder block in order to record an acoustic emission. The detailed description of experimental setup and conditions are out of scope in this paper and can be found in [13, 17]. Thus, for the purpose of this research, the available database of the engine performance data is used.

Table 1. Test engine specification

Definition	Dimension	Value
Engine Type	[--]	Yanmar AYG20L
Cylinders	[--]	6 (in-line)
Bore/Stroke	[mm/mm]	155/180
Power	[kW]	434
Speed	[rpm]	1800/1500
BMEP	[bar]	14

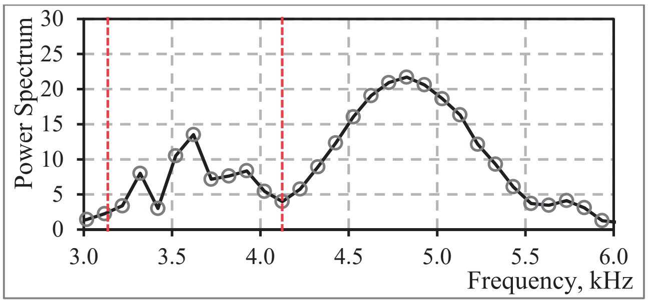
2.2. Characteristics of Knock

The knock combustion phenomenon is caused by the spontaneous ignition spots in the volume of the air-fuel mixture during the end of compression or/and combustion cycles. The shock waves, produced as the result of a rapid release of the chemical energy, propagate through the volume and excite structure vibration, combustion chamber acoustic resonance, and lead to a high-frequency oscillation in-cylinder pressure. The frequencies of transverse vibration modes in a circular cylinder can be estimated from the sound wave theory as found in [14]. For the given engine type and rated in-cylinder temperature, the resonant frequency is in the range of 3.2 ~ 4.2 kHz. For the tests engine, the typical frequency components of the in-cylinder pressure signal obtained with the Fourier analysis are depicted in Fig. 1. Figure 1a corresponds to the normal combustion cycle where the spectrum components of around 5 kHz are dominant and specific to the test engine, as discussed in [13]. On the contrary, Fig. 1b illustrates the combustion cycle with some sign of knocking where the spectrum components in the frequency range of 3.2 – 4.2 kHz are dominant. Thus, the combustion cycle pattern consists of spectrum components of different power related to the knocking and non-knocking operation.

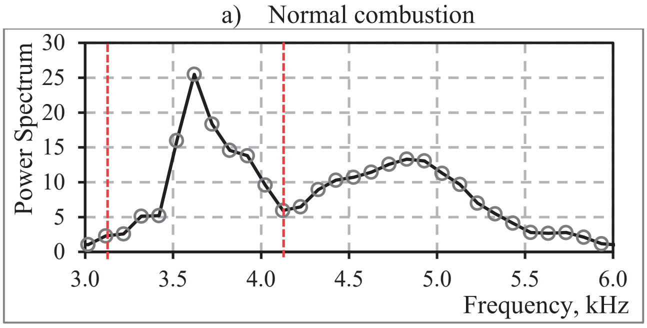
Many researchers investigated the properties of the in-cylinder pressure signal in view of knock onset, and the principal conclusions are briefly summarised below:

- Knock appears over a short combustion period;
- Interaction between the front flame and spontaneous ignition spots, in general, affects the knock amplitudes;
- Cycle-to-cycle variability: the typical knock oscillations appear sporadically, with varying levels (as also seen in Fig. 1);
- Resonant frequency changes with the piston position, combustion temperature and composition of the air-fuel mixture.

The last two points imply the nonstationary and stochastic nature of the knock event and justify specific difficulties in knock detection [5, 6].



a)　Normal combustion



b)　Sign of knocking

Fig. 1 Typical frequency pattern of in-cylinder pressure of the test engine

2.3　Knock features extraction

It is undeniable that the varying spatial[1] frequency properties of the knocking phenomena are hardly interpretable by the Fourier analysis, moreover buried in the background noise (that is prominent for the vibration and acoustic emission (AE) signals). Another way of looking at knock features is to apply a time-frequency analysis. Wavelet transform (WT) is able to localise both time and frequency components. The Morlet wavelet, as a basis function of WT, is especially suitable for the knock features analysis, owing to its analogy with the Fourier transform [15]. The WT defines a joint time-frequency density map called a spectrogram or wavelet power spectrum. In this way, the high energy components can be easily localised in time and frequency domains allowing for discrimination of the knock features.

The joint energy density time-frequency distribution provided by the WT can further be subjected to singular value decomposition (SVD). The SVD is the general technique used in signal processing for noise thresholding, detection of subspace with the highest energy density, and information retrieval. Although the SVD processing provides a suitable tool for features extraction, it may span the entire range of time and frequency with no single energy concentration at any given time or frequency. To overcome above deficiency, a modification to the SVD was proposed [16] and successfully applied for the knock features extraction [17]. The transformation applied to the SVD results in the rotation of original singular vectors into a set of principal axes concentrating the densities and minimising the mutual energy overflow between the two density functions. Figure 2 illustrates such transformation of a signal, and Fig.3 visualises the results of transformation for the two different engine cycles: the one where the spectrum components of normal operation are dominant and the other where the spectrum components of knock are dominant. The transformed singular value decomposition (TSVD) effectively highlights the essential features with the highest energy density and reduces the effects of noise from the original wavelet spectrogram.

Feature extraction from the joint time-frequency distribution provided by the WT is based on decoupling it into discrete density functions for the purpose of moments calculation [18]. Using the TSVD, the spectrogram matrix $\mathbf{P}_{WT}$ can be decomposed into a sum of basis matrices weighted by a singular value. Where every basis matrix is a discrete density function due to the orthonormality of the original singular vectors composing the matrices **Y** and **X:**

$$\widetilde{\mathbf{P}}_{WT} = \mathbf{Y}\,\mathbf{Z}\,\mathbf{X}^T = \sum_{i,j} z_{i,j} \left\{\widetilde{\mathbf{P}}_{WT}\right\}_{i,j}, \quad \because \left\{\widetilde{\mathbf{P}}_{WT}\right\}_{i,j} = \mathbf{y}_i \mathbf{x}_j^T \qquad (1)$$

Thus, a subset of temporal and spectral moments can be obtained:

[1] With respect to the engine, spatial means the angle position of the crankshaft which is also can be converted to time.

$$\begin{aligned} \{t^p\}_i &= \sum_{k=1}^{m} (t_k)^p \,\tilde{\mathbf{y}}_i(k), \quad \{f^q\}_j = \sum_{l=1}^{n} (f_l)^q \,\tilde{\mathbf{x}}_j(l) \\ &\because \tilde{\mathbf{y}}_i(k) = (y_{k,i})^2 \qquad \because \tilde{\mathbf{x}}_j(l) = (x_{l,j})^2 \end{aligned} \tag{2}$$

The first moments $(p = q = 1)$ estimate time-correlated instantaneous frequencies, and the second moments $(p = q = 2)$ estimate the time-span and frequency bandwidth for each feature in the distribution. Then, each extracted feature can be defined as a five elements vector:

$$\begin{aligned} F_i &= \left(\tilde{z}_{i,j}, \bar{t}_i, \bar{f}_j, \hat{t}_i, \hat{f}_j\right) \\ &= \left\{ \frac{z_{i,j}^2}{\max(z_{i,j})^2}, \langle t \rangle_i, \langle f \rangle_j, \sqrt{\langle t^2 \rangle_i - \bar{t}_i^2}, \sqrt{\langle f^2 \rangle_i - \bar{f}_i^2} \right\} \end{aligned} \tag{3}$$

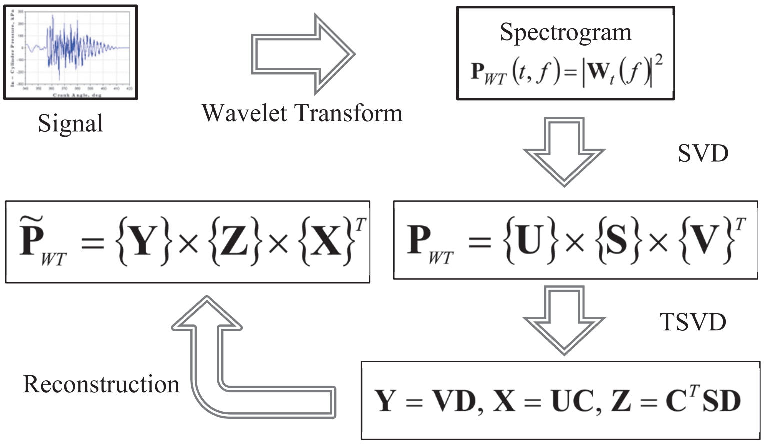


Fig. 2 Illustration of signal processing using WT and TSVD

The number of features derived from every signal instance is determined by the significant energy content, represented by the cumulated percentage of the sorted singular values $z_{i,j}^2$. Usually, a more than 70% threshold is the recommended value of choice. Table 2 demonstrates the first five features extracted from the signal instance. Repeating the feature extraction procedure for a large number of consecutive cycles, a map of features characterising the state of an engine can be obtained, as shown in Fig. 4. The features map demonstrates only the first moments. Also, the marginal density functions of the individual component are jointly illustrated.

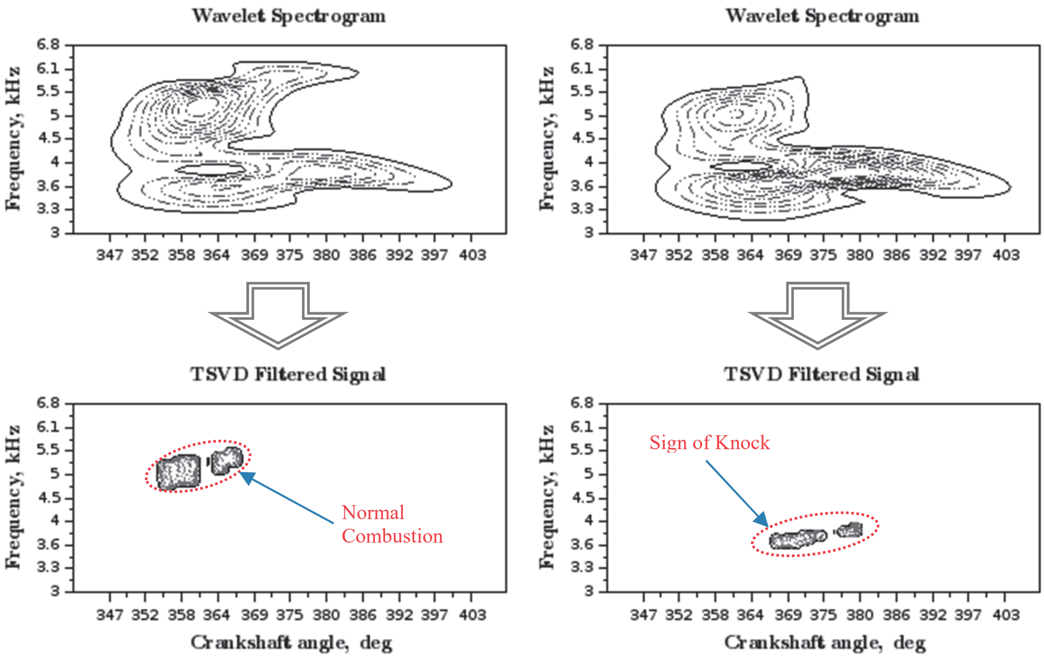


Fig. 3 Knock features highlight with the Wavelet-TSVD signal processing

As can be observed from the features map, presented in Fig.4, the first cluster of features related to the wide frequency span 4.0~5.5 kHz, and concentrated around the top dead centre (TDC) position of the crankshaft (360 deg) appears most frequently, and these features are solely related to the normal operation of a given test engine. On the other hand, appearing sporadically knock-related features form a separate cluster, spanned over broad crankshaft angles (360 ~ 390 deg) and concentrated at a narrow frequencies band of 3.2~3.8 kHz. The shared densities of the features are then redistributed between the clusters in case of varying knock intensity. From this perspective, it becomes evident that the clustering and classification algorithms of ML are highly suitable for the task of the knock intensity identification and subsequent chapters elaborate on the performance of two classification algorithms: SVM and SOM.

Table 2. Example of features extracted from the signal instance

Item	$\tilde{z}_{i,j}$	$\bar{t}_i$, deg	$\bar{f}_i$, kHz	$\hat{t}_i$, deg	$\hat{f}_i$, kHz
1	0.77	358	3.22	8.6	0.0
2	0.84	358	3.30	8.6	0.0
3	0.77	366	3.30	2.9	0.0
4	0.77	367	3.30	5.4	0.0
5	0.75	355	3.34	7.0	0.0

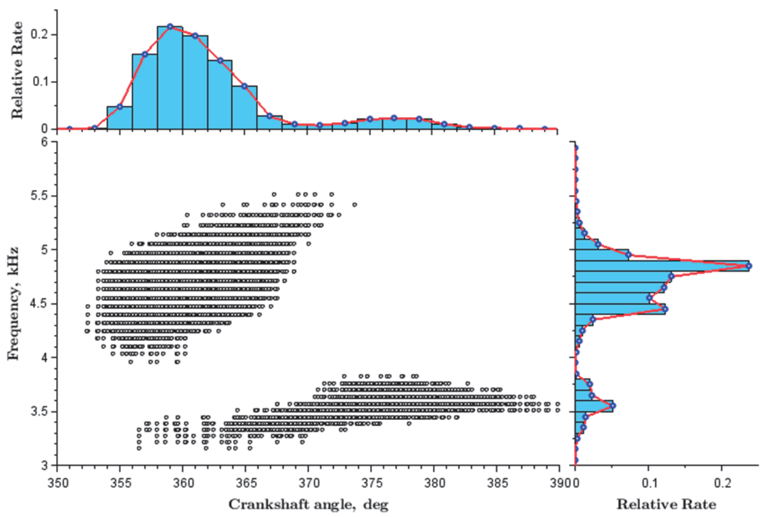


Fig. 4. Features distribution in 2000 consecutive cycles of the normally operated engine

3. Methods of Machine Learning for the Knock Classification

3.1 Application of Support Vector Machine (SVM)

The SVM aims to classify input data into one of the available clusters providing that a features mapping model exists. In this respect, at the learning stage, the SVM constructs a hyperplane that separates the features space into two or more clusters, and by maximising the margins, the risk of misclassification is minimised. This means that the separating hyperplane has a maximal distance to the points at cluster boundaries [9]. The nonlinearity is introduced by a kernel function which in most cases is the Radial-Basis Function (RBF). The SVM belongs to the class of supervised learning algorithms in which a model is built on the set of labelled data. However, the provided engine features map has no labels, and thus these have to be assigned by a clustering algorithm beforehand.

Clustering is the task of grouping a set of input data in such a way that the data in the same cluster are more similar to each other than to those in other clusters. In the Fuzzy c-means (FCM) clustering algorithm [19], the degree of similarity is defined by introducing the membership to clusters. The latter is called a fuzzification of the cluster configuration, and in many situations, FCM clustering is more natural than hard clustering. Moreover, the degree of membership allows to get rid of outliers and defines the robust edge of the cluster.

Let's assume a set of n variables $X = \{x_1, x_2,\dots, x_n\}$ where x_i is a d-dimensional point. The fuzzy clustering is a collection of k clusters prototypes, $C_1, C_2,\dots, C_k$ and a partition matrix $W = w_{i,j} \in [0, 1]$, for $i = 1\dots n$ and $j = 1\dots k$, where each element $w_{i,j}$ is the weight that represents the degree of membership of data point i to cluster C_j. The clustering criterion is associated with the generalised least-squared errors objective function in the following form:

$$J_m(C,v) = \sum_{i=1}^{n}\sum_{j=1}^{k}\left(w_{i,j}\right)^m d_{i,j}^2 \tag{4}$$

where $v = \{v_1, v_2,\dots, v_k\}$ is the vector of centers and $v_i = \{v_{i1}, v_{i2},\dots, v_{in}\}$ is the center of cluster i. Weighting exponent m controls the fuzziness of the clusters.

The Euclidean distance metric is often used to measure the distance from input data point $\boldsymbol{x}$ and the i^{th} cluster prototype:

$$d_{i,j}^2 = \left\|x_i - v_j\right\|_A^2 = \left(x_i - v_j\right)A\left(x_i - v_j\right)^T \tag{5}$$

where A_i is the diagonal matrix. In the modified FCM by Gustafson and Kessel [20], the matrices A_i are also the decision variables and the size of $\left|A_i\right|$ is constrained to a specific value.

As an initial step, the algorithm requires assumptions on the number of clusters in the data, and obviously, there are only two clusters: the one is the dense features defining regular engine operation, and the other is the sparse features related to knock. The results of the FCM clustering applied to the engine features map are illustrated in Fig. 5, followed by the clusters decision boundaries of the SVM model after learning, depicted in Fig. 6. Here it should be noted that the data for ML reduced to the standard normalised scale. Thus '0' point corresponds to the following actual coordinates [360, 3600].

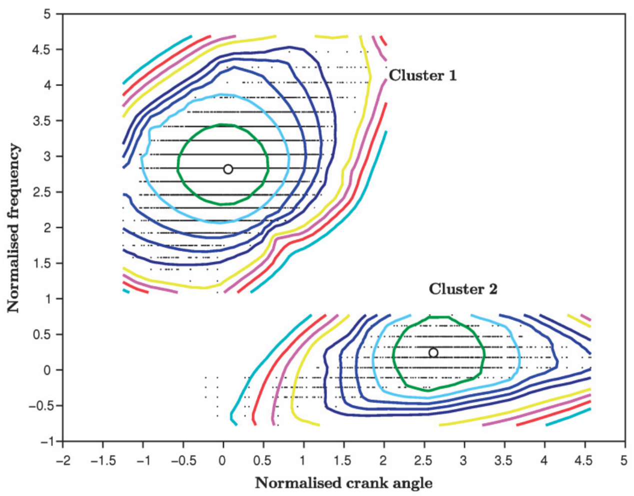


Fig. 5. FCM Clustering of the features map. Iso-lines depict the regions with a membership degree larger than 60%

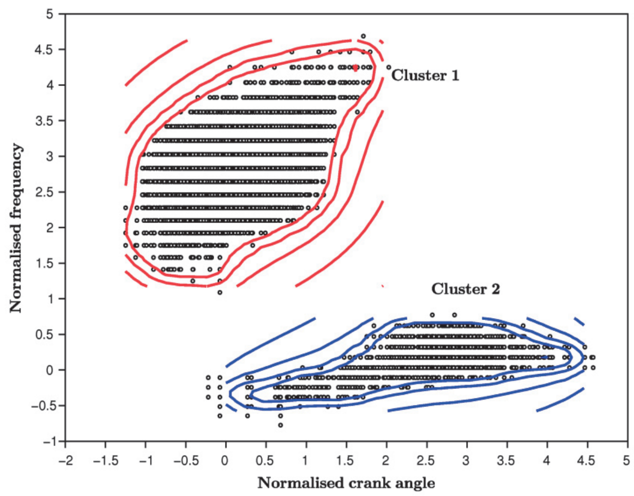


Fig. 6. Clusters decision boundaries of the SVM model

3.1.1　SVM classification results

The built SVM model now is able to classify the data. New data samples are mapped into the model space and predicted to belong to one of the clusters. In regular operation, most of the data belong to cluster 1, but the more the engine knocking, most of the data appear in cluster 2. A straightforward idea of knock intensity supervision is to estimate the probability that the data from several consecutive cycles belong either to cluster 1 or 2 by using the following ratio:

$$P(i)=\frac{\#\left(\mathbf{x}_n \in V_i\right)}{\#\left(\mathbf{x}_n\right)}, n=1,\ldots,k \tag{6}$$

where V_i is the decision boundaries for cluster i, and k is the number of samples. The symbol # denotes the number of samples.

The data were collected for the different intensities of knocking, which were initiated artificially by changing the engine parameters. For every running condition, 1000~2000 consecutive cycles were recorded, and features were extracted sequentially for every 10 cycles and then were classified for knock intensity which results are depicted in Fig. 7. The knocking index shows the probability of data hit for cluster 2 and vice versa. Figure 8 depicts the results of knock index calculation based on band-pass filtering of in-cylinder pressure and averaging over all collected cycles for the given running conditions (this is the standard method of knock intensity estimation). Such results confirm the intensity knocking classification by means of the SVM, with the only difference that the ML technique is able to supervise the intensity of knock continuously based on a few consecutive cycles.

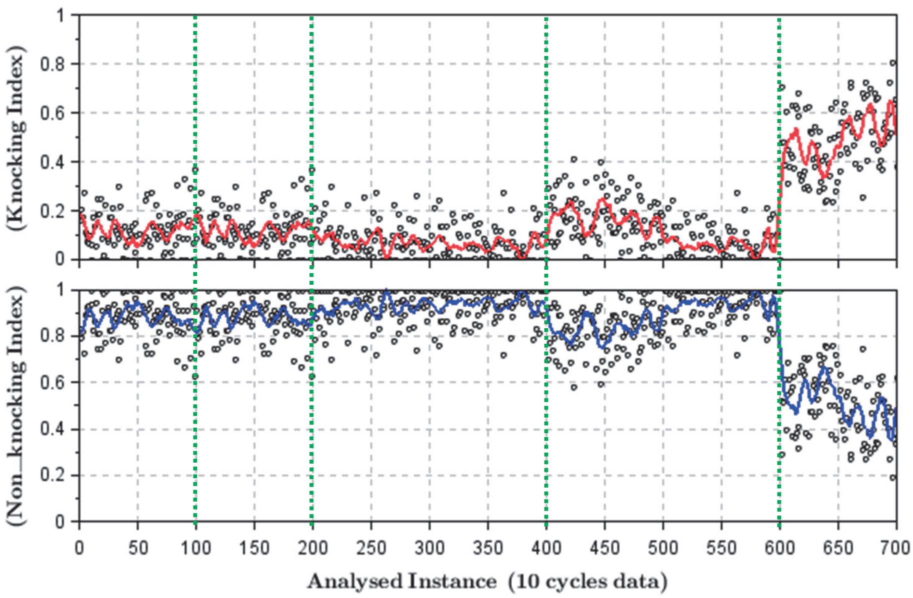


Fig. 7. Knock intensity classification with the SVM

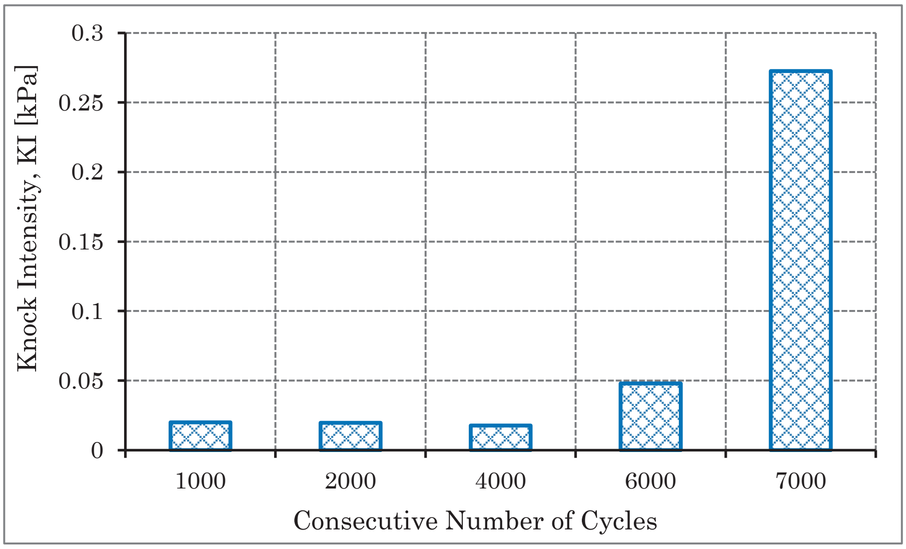


Fig. 8. Knock intensity at varying engine running conditions

3.2 Application of Self-Organising Map (SOM)

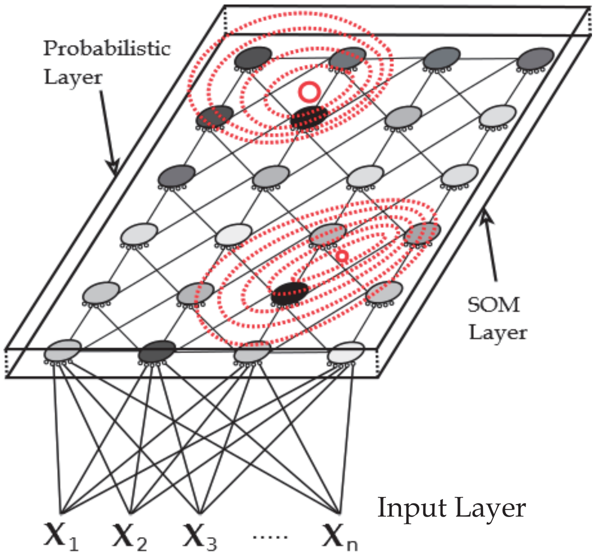
On the contrary to supervised learning provided by SVM, the Self-Organised Map fits with the unsupervised learning of the underlying data topology. The SOM is a nonlinear projection method where a high-dimensional input space is mapped into a lower-dimension (usually two dimensional) regular lattice of neurons. In this way, the SOM preserves the topology of input space – feature vectors close to each other in input space will be mapped correspondingly by the SOM neurons, thus preserving the clusters separation. The structure of SOM is that of a neural network with an excitation input layer and a two-dimensional array of *M*x*N* neurons in the output layer. The exact formulation of SOM is not elaborated in this paper.

Concerning the problem of knock intensity classification, a new vector of features samples has to be compared with the model vectors of the SOM, but the output layer does not define any probability measures to determine the level of similarity [21]. The problem can be solved by adding the additional output layer, which is based on the FCM clustering. The outline of the hybrid SOM is depicted in Fig. 9.

Owing to the fact that the structure of the cluster from the input space is preserved by the low-dimension topology of SOM, the application of the FCM clustering algorithm is straightforward, providing an estimation of the centres of the cluster.

For the decision on knock intensity, a response function of the distance between the input feature vector and cluster centres is proposed in the following form:

$$r_i = e^{-d_i}, \quad d_i = \sum_{j=1}^{n} \left(x_j - v_i \right)^2, \quad \because i = 1,2 \tag{7}$$



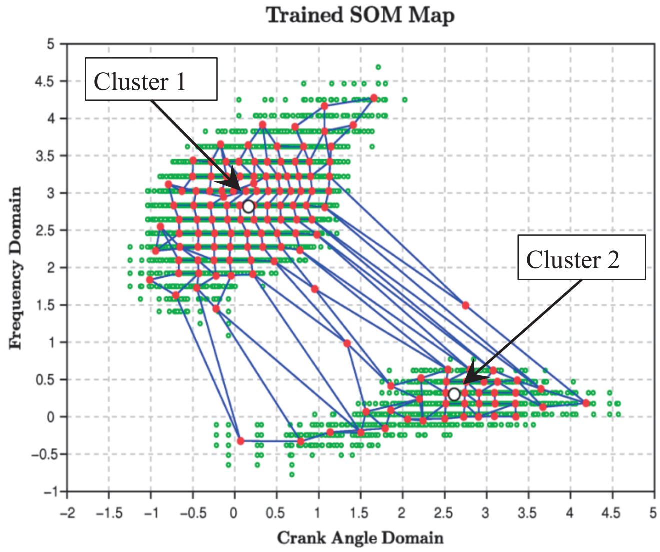


Fig. 9. The topology of Probabilistic SOM classifier　　Fig. 10. The input features space mapped by the SOM

The output of the response function can be considered as a "fuzzy indicator" of the input feature hit for cluster i, since $r_i \in [0,1]$. Finally, the knock intensity is estimated by measuring the frequency of input features classified to belong to either of the clusters, utilising Eq. (6).

3.2.1　SOM classification results

The SOM with the layer size 12x12 neurons was used to map the topology of input features space, as illustrated in Fig. 10. In addition, clusters centres, identified by FCM clustering algorithm, are indicated by bold dots. After that, knock intensity classification is straightforward and similar to that with SVM application. Results are demonstrated in Fig. 11, and as evident, are similar to the latter case, as well as correlate reasonably well with the standard method of knock intensity classification reported in Fig. 8.

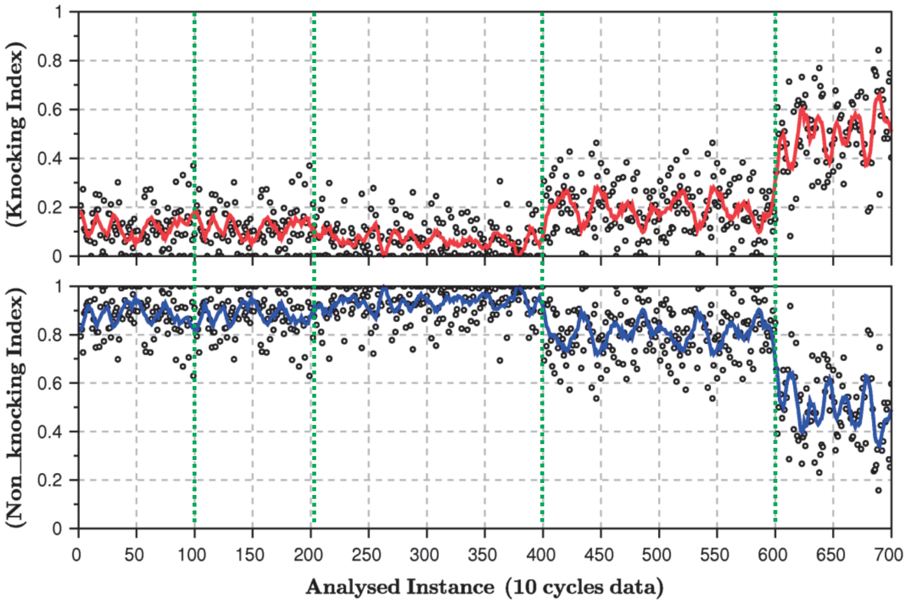


Fig. 11. Knock intensity classification with the Hybrid-SOM

3.3 Application of ML technique for knock detection from acoustic emission data

The ML methods of knock intensity monitoring, described in the preceding chapters, were applied to the measured in-cylinder pressure traces, which directly contain the components of knock combustion, and providing visible separation of features to the clusters. However, in-cylinder pressure monitoring inevitably increases the cost and complexity of the engine design during production, and therefore a direct measurement of in-cylinder pressure is rarely available in operation. On the other hand, oscillating pressure due to knock combustion induces acoustic emission that propagates outside of the engine and thus can be sensed by the non-intrusive methods. Thus, the availability of an additional measurement channel for knock intensity monitoring increases the reliability of engine safe operation.

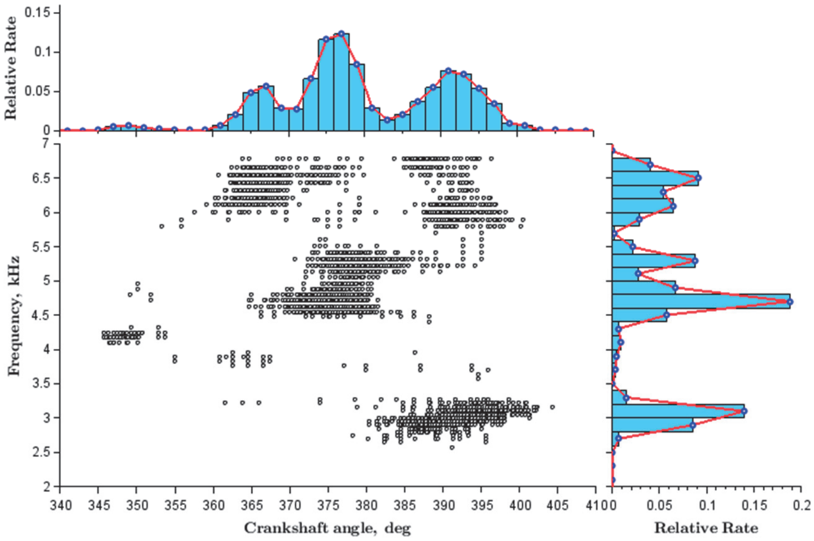


Fig. 12. Features extracted from AE signal of 2000 consecutive cycles of the normally operated engine

For the purpose of this research, the AE was recorded by a microphone installed in close proximity to the instrumented cylinder block. The gathered data of a number of consecutive cycles were then processed by the wavelet analysis followed by the TSVD analysis. The distribution of features is illustrated in Fig. 12. Although the distribution shows a certain degree of similitude with the features extracted from the in-cylinder pressure signal (Fig. 4), there is also a high share of uncertain features on the map. The latter, perhaps, comes from the AE of the adjacent cylinders. In that respect, the direct application of clustering and classification methods is not applicable because the uneven distribution of data does not provide a clear configuration of clusters. Thus, certain modifications of algorithms are required, as explained below.

3.3.1 Adaptation of ML technique for acoustic emission data processing

For the sake of brevity, only the application of hybrid (probabilistic) SOM is elaborated hereinafter. As was mentioned above, the features extracted from the AE data form uneven density distribution with the uncertain features on the map, which also form separate clusters. At the same time, the distribution of neurons of the SOM is characterised by even density. In order to match the input space density distribution to the SOM neurons space, the convex combination method is applied. The method consists in simultaneous and gradual adaptation of input space vectors and neuron weights space during SOM unsupervised learning. That is done by the modification of a set of n input vectors $\{X\}^{n} = \{x_1, x_2,\ldots, x_n\}$ as follows:

$$\left\{\hat{X}\right\}^{n} = \beta\left(k\right)\left\{X\right\}^{n} + \frac{1-\beta\left(k\right)}{\sqrt{n}}, \quad \therefore \beta\left(k\right) = \frac{e^{k} - e^{-k}}{e^{k} + e^{-k}} \tag{8}$$

where $\beta(k)$ is the monotonically increasing function with the domain between -1 and 1, $k \in [0,1]$ is the normalised iteration step.

At the commence of learning $\beta(k) = 0$ and the input space vectors are concentrated at the point with the coordinates $\{1/\sqrt{n}\}$. Similarly, the weights of the neurons are also placed at the same point. As the number of iteration step increases, the input space vectors diverge from the point towards their nominal values enabling co-joining the weights of neurons with the data and thus preserving density distribution of data.

The SOM with the layer size 12x12 neurons was used to map the topology of input features space, and Figure 13 illustrates the structure of the input space cluster preserved by the topology of SOM. Although most neurons represent the required clusters (regular operation and knock combustion), uncertain features also form stray clusters.

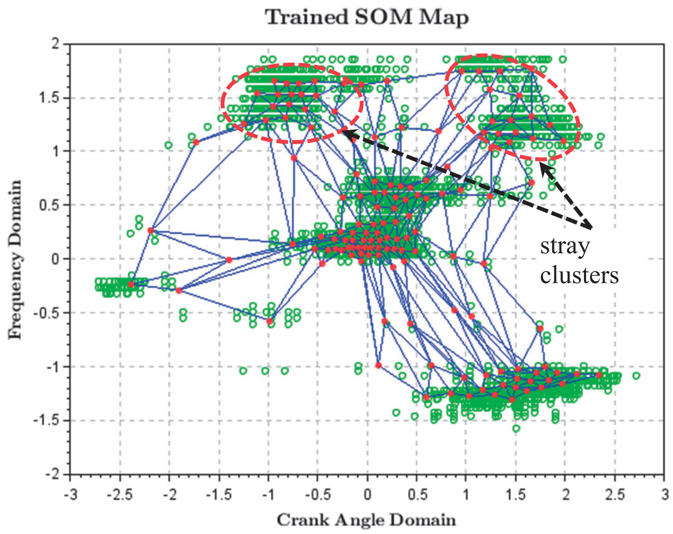


Fig. 13. The topology of input features space preserved by the SOM neurons

As a result, the FCM clustering algorithm also captures the centres of stray clusters, as can be seen in Fig. 14a. The remedy to this problem is a cluster merging algorithm. The idea is to merge the clusters based on the similarity measure between them. The clusters similarity matrix [22] can be defined as follows:

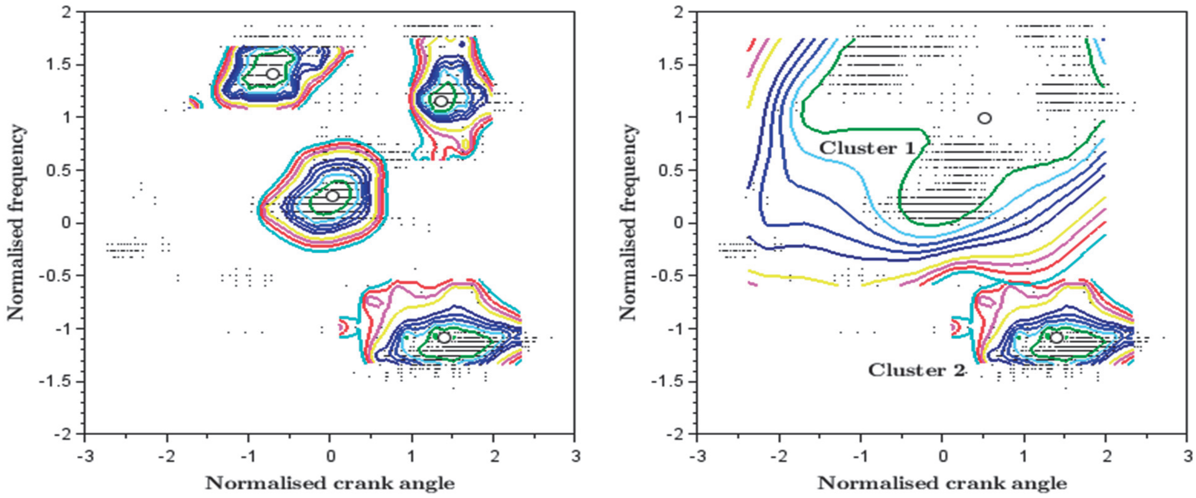
$$FR_{ij} = \frac{dp_i + dp_j}{d\upsilon_{ij}} \tag{9}$$

where $d\upsilon_{ij}$ denotes the dissimilarity between two clusters C_i and C_j, i.e. Euclidian distance between clusters centres; In turn, a fuzzy dispersion dp_i is the fuzzy measure of the radius of cluster C_i, defined as follows:

$$dp_i = \sqrt{\tfrac{1}{n}\sum_{\boldsymbol{X}\in C_i} \boldsymbol{\mu}_i^m D_M^2\left(\boldsymbol{X}\right)}, \quad \therefore D_M\left(\boldsymbol{X}\right) = \sqrt{\left(\boldsymbol{X}-\boldsymbol{V}_i\right)^T S_i^{-1}\left(\boldsymbol{X}-\boldsymbol{V}_i\right)} \tag{10}$$

where D_M $(\boldsymbol{X})$ is the Mahalanobis distance, $\boldsymbol{S}_i$ is the cluster covariance matrix (available as the output of FCM algorithm), and $\boldsymbol{V}_i$ is the cluster prototype (centre).

Two clusters are considered similar if the corresponding FR_{ij} is the maximum and thus can be merged. The cluster merging process repeats until only two clusters remain, as illustrated in Fig. 14b.



a) Original clustering with stray clusters b) Result of merging process

Fig. 14. FCM Clustering of the SOM map. Iso-lines depict the regions with a membership degree larger than 60%

3.3.2 Results of knock classification in the AE data

After obtaining the clear separation of features space into two clusters, the knock classification is straightforward and similar to that of the past application (Eq. (6)). Results are demonstrated in Fig. 15. The data sets for the different intensities of knocking are designated by the vertical dashed lines. Because of the presence of uncertain features in the data, which also resulted in broad decision boundaries of cluster 1, the knock index is estimated for 100 consecutive cycles. Although the results are not as clear as in the preceding cases, the tendency correlates reasonably well. Thus, set 1 and set 2 correspond to non-knocking condition; set 3 indicates a jump in knocking condition, and this can be related to contribution from the adjacent cylinders, states of combustion in which are unknown; set 4 indicates an increase of knock index at the beginning followed by the relaxation, the same as in-cylinder pressure features indicate; finally set 5 shows gradually increasing knock index.

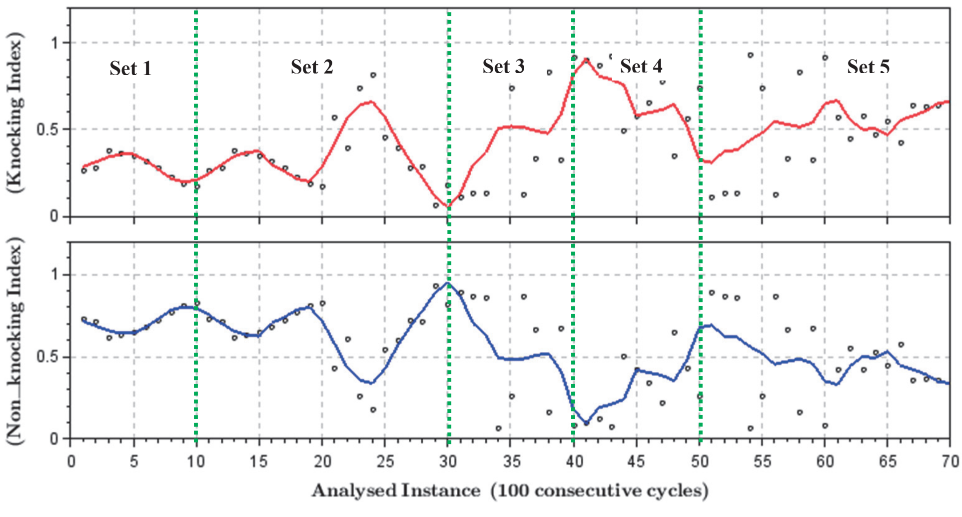


Fig. 15. Knock classification in the AE data with the Hybrid-SOM

4. Conclusions

In this study, the problem of knock intensity monitoring was addressed, and the methods from the field of machine learning have been utilised successfully. The performance of the ML algorithms depends, to some extent, on the features extracted from the raw data and containing useful information. Thus, the developed feature engineering algorithm justified the applicability of the two popular classification algorithms, notably SVM and SOM. However, nor of them are applicable standalone to quantify the knock intensity. Only the combination with the other ML algorithms can bring the fruits.

The supervised learning of the SVM model requires the data to be labelled, and in this respect, the fuzzy c-means clustering is built on top of the SVM. On the other hand, owing to the two main properties of SOM, which are dimension reduction and topology preservation, the clusters separation of the input features space was mapped by the limited number of the model neurons in an unsupervised manner. Finally, in order to quantify the similarity between the features characterising the present engine running condition and the reference one, a probabilistic decision rule has been added to the output layer of each model. Thereby, the intensity of knocking is characterised as a deviation from the reference operating condition. The summary of the developed algorithms of knock intensity monitoring is depicted in form of block diagram in Fig. 16.

Although the processed in-cylinder pressure signal provides a visible separation of features to the clusters ensuring robust classification of knock combustion severity, the in-cylinder pressure measurements are not available on the production engines. In this respect, the non-intrusive method of acoustic emission processing has been scrutinised closely for the application of the ML technique. However, features distribution that appeared in the AE data is buried with the stray features. Specific modification introduced into the SOM training algorithm and cluster merging improved to some extent the appearance of the clusters. However, the classification still suffers from the stray features coming from the adjacent cylinders. The latter requires further elaboration.

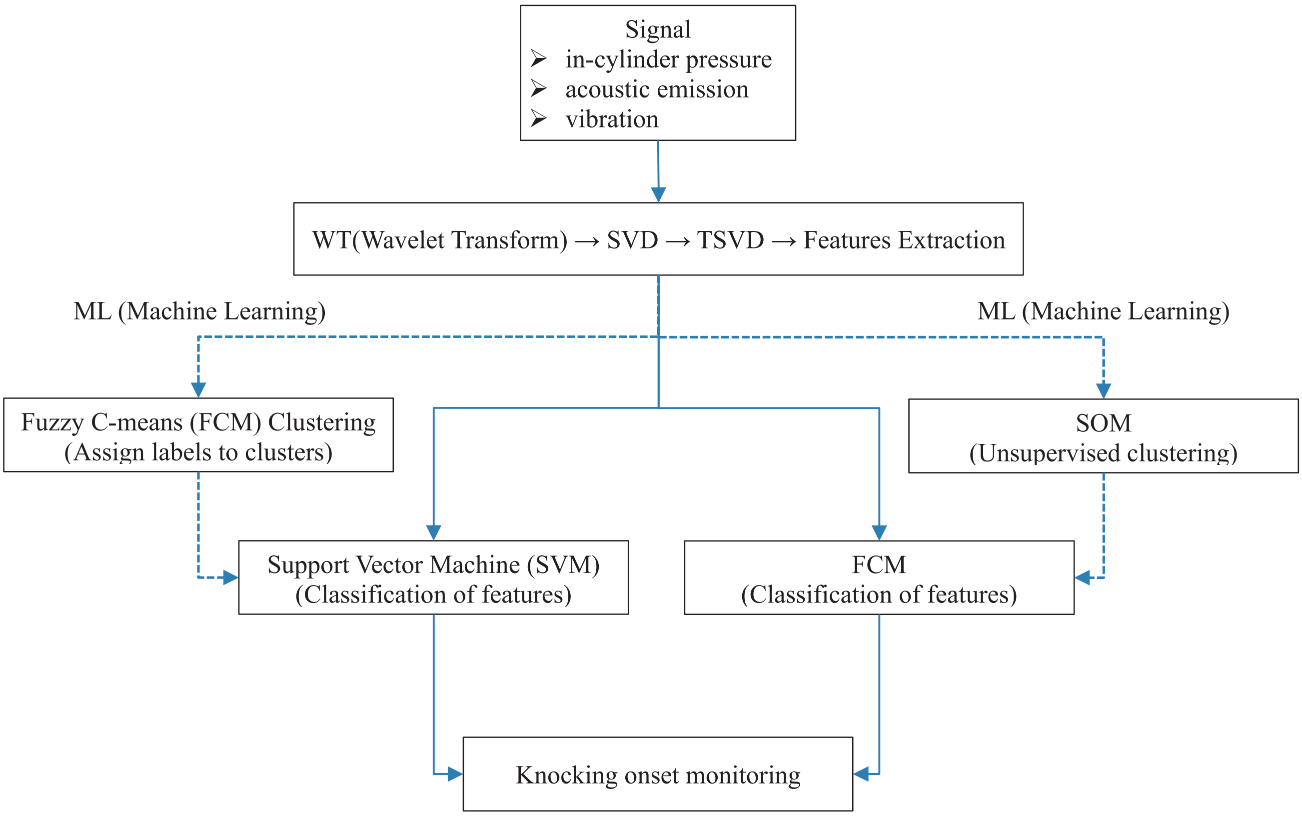


Fig. 16 The block diagram of knock intensity monitoring algorithms

References

1) Healthy freight rates and 2018 LNG shipping predictions, https://www.lngworldshipping.com/news/view,healthy-freight-rates-and-2018-lng-shipping-predictions_50323.htm, last accessed 2019/05/31
2) Cavina N., Corti E., Minelli G., Moro D., and Solieri L.: Knock Indexes Normalization Methodologies, SAE Technical Paper 2006-01-2998.
3) Ker S., Bonnardot F., Duval L.: Algorithm Comparison for Real Time Knock Detection. In: International Conference on Acoustic, Speech, and Signal Processing, ICASSP '07, pp.397-400, (2007).
4) Millo F., Lavarino G., Cafari A.: Knock in Dual Fuel Engines: a Comparison Between Different Techniques for Detection and Control. CIMAC Congress, Paper No.:212, (2010).
5) Akimoto K., Komatsu H., Kurauchi A.: Construction of Knock Detection Logic by Pattern Recognition Using Short-Time Fourier Transform. In: Honda R&D Technical Review, (2012).
6) Duval L., Corge G., Tran V.B., and Leduc P.: Noise Robust Spark Ignition Engine Knock Detection with Redundant Wavelet Transform. In: International Conference on Noise and Vibration, (2002).
7) Yong H.Z., Yan J., Chen Y.: Study of Engine Noise Based on Independent Component Analysis. In: Journal of Zhjiang University, vol.8, issue.5, (2007).
8) Lowe D.P., Lin T.R.: Diesel Knock Combustion and its Detection Using Acoustic Emission. In: Journal of Acoustic Emission, vol.29, (2011).
9) Rychetsky M., Ortmann S., Glesner M.: Support vector approaches for engine knock detection, In: International Joint Conference on Neural Networks, IJCNN '99, vol.2, pp.969-974 (1999).
10) Lazarescu D., Lazarescu V., Ungureanu M.: Knock Detection Based on SOM, In: 7th Seminar on Neural Network Applications in Electrical Engineering, NEUREL 2004, pp.117-120 (2004).
11) Lopez-Rubio E.: Probabilistic Self-Organising Maps for Continuous Data. In: IEEE Transaction on Neural Networks, vol.21, issue 10, pp.1543-1554 (2010).
12) Corchado E., Grana M., Wozniak M.: New trends and applications on hybrid artificial intelligence systems. Neurocomputing, vol.75, issue 1, pp.61-63 (2012).
13) Ichikawa Y., Sekiguchi H., Bondarenko O., Yoo D.-H., and Hirata K.: Effects of Gas Composition on Combustion, Emission and Knocking Characteristics of Marine Lean Burn Gas Engine. In: ASME Internal Combustion Engine Fall Technical Conference, (2016).
14) Draper C.S.: The Physical Effects of Detonation in a Closed Cylindrical Chamber, NACA, Report No.:493, pp.361-379, (1933).
15) Huang M.-C.: Wave Parameters and Functions in Wavelet Analysis. In: Journal of Ocean Engineering, vol.31, (2004).
16) Groutage D., Bennink D.: A New Matrix Decomposition Based on Transforming the Basis Sets of the Singular Value Decomposition Yields Principal Features for Time-Frequency Distributions. In: Advanced Signal Processing Algorithms, Architecture and Implementations, vol.4116, (2000).
17) Bondarenko O., Fukuda T.: Potential of Acoustic Emission in Unsupervised Monitoring of Gas-Fueled Engines. In: 10th IFAC Conference on Control Applications in Marine Systems, (2016).
18) Guo Y., Kareem A.: System Identification through Nonstationary Response: Wavelet and Transformed Singular Value Decomposition - Based Approach. In: Journal of Engineering Mechanics, vol.141, issue.7, (2015).
19) Bezdek J.C., Ehrlich R., Full W.: FCM: The Fuzzy c-Means Clustering Algorithm. In: Computer & Geosciences, vol.10, issue.2-3, pp.191-203, (1984).
20) Gustafson D.E., Kessel, W.C.: Fuzzy clustering with a covariance matrix. In: Proc. IEEE CDC, vol. 2, pp.761-766 (1979).
21) Alhoniemi E., Himberg J., Vesanto J.: Probabilistic measures for responses of Self-Organizing Map units, In: Proceedings of the International ICSC Congress on Computational Intelligence Methods and Applications, CIMA '99, pp.286-290, (1999).
22) Xiong X., Chan K.L., Tan K.L.: Similarity-Driven Cluster Merging Method for Unsupervised Fuzzy Clustering, In: Proc. of the 20th Conference on Uncertainty in Artificial Intelligence (UAI2004), pp. 611-618 (2004).