SPECTRUM CLASSIFICATION FOR EARLY FAULT DIAGNOSIS OF THE LP GAS PRESSURE REGULATOR BASED ON THE KULLBACK-LEIBLER KERNEL

Tsukasa Ishigaki[1], *Tomoyuki Higuchi*[2] *and Kajiro Watanabe*[3]

[1]Department of Statistical Science, The Graduate University for Advanced Studies and JST CREST
4-6-7 Minami Azabu, Minato-ku, Tokyo 106-8569, Japan
[2]Institute of Statistical Mathematics and JST CREST
4-6-7 Minami Azabu, Minato-ku, Tokyo 106-8569, Japan
[3]Department of System Control Engineering, Hosei University
3-7-2 Kajinocho, Koganei-shi, Tokyo 184-8584, Japan
ishigaki@ism.ac.jp; higuchi@ism.ac.jp; bob@k.hosei.ac.jp

ABSTRACT

The present paper describes a frequency spectrum classification method for fault diagnosis of the LP gas pressure regulator using Support Vector Machines. Conventional diagnosis methods are not efficient because of problems such as significant noise and nonlinearity of the detection mechanism. In order to solve these problems, a machine learning method with the Kullback-Leibler (KL) kernel based on the KL divergence is introduced into spectrum classification. We use the normalized frequency spectrum directly as input with the KL kernel. The proposed method demonstrates a higher accuracy than popular kernels, such as polynomial or Gaussian kernels, or the conventional fault diagnosis method and Gaussian Mixture Model with the KL kernel for the examined problem. The high classification performance is achieved by using an inexpensive sensor system and the machine learning method. This method is widely applicable to other spectrum classification applications without limitation on the generality if the spectrums are normalized.

1. INTRODUCTION

Development of a learning machine for classification using kernel methods, as typified by the Support Vector Machine (SVM), has influenced a number of fields. The strategy of classification before the introduction of kernel methods and machine learning was that given high-dimensional data was mapped onto a low-dimensional space that characterizes the properties of the data well. Then, a number of useful parameters for classification are extracted and optimized statistically. Conversely, the strategy of kernel methods is to generate a map from given high-dimensional data onto a space of higher dimension and then create a linear classifier in the space. Kernel methods have been applied successfully in various fields, including image, text and speech classification, data mining and bioinformatics [1, 2]. Recently, kernel methods have been investigated for application to fault diagnosis [3, 4].

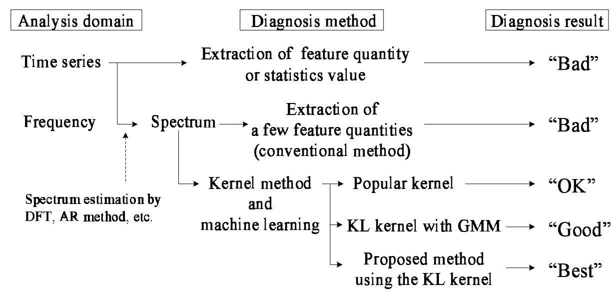

Fig. 1. Outline of diagnosis methods considered herein

Fault diagnosis must be performed in order to ensure the safety of machines and instruments. Diagnosis methods based on time series, spectrum or spatial analyses have been proven useful, and a feasible method is generally chosen from among these methods based on the particular situation. The spectrum analysis method is generally used for the diagnosis of machines that have a vibrating or rotating structure. The use of spectrum analysis in this case is appropriate [5]. The analysis method in a time domain is not suitable in this case because machines that have a vibrating or rotating structure often require inspection of the steady state vibration or rotation rather than inspection of the transient response. In such cases, the spectrum diagnosis method is superior to time-series-based diagnosis methods. Conventional spectrum analysis methods for fault diagnosis, e.g., the half-power method, extract a small number of feature quantities from spectrum peaks caused by intended modes, and these feature quantities are then investigated. Before the advent of machine learning methods, this was the only the classification strategy. The above methods are effective if the system model of the targets or the properties of observation data are well known. However, if the model of the targets is unknown or the observed data is nonlinear, it is difficult to apply the conventional diagnosis method because the extraction of the feature quantities is often not correct. Therefore, fault diagnosis using machine learning methods, such as kernel methods, in which the entire set of spectrum information is used as a feature, is a natural development.

Here, we consider the problem of fault diagnosis of LP gas equipment. LP gas is a significant energy source that is used in various applications in daily life. The use of LP gas is not safe unless the LP gas supplying equipment is properly maintained. The LP gas pressure regulator (hereinafter called the regulator) is one piece of equipment used in LP gas supply. The regulator regulates the pressure by the movement of a rubber diaphragm fixed in the regulator. In rare cases, it has been reported that the rubber diaphragm will begin to vibrate abnormally even though the lifetime of the diaphragm has not yet been reached. Abnormal vibration of the diaphragm can have the same effect as accelerated deterioration of the rubber material, and a gas leak or fire may occur if the diaphragm bursts. However, direct observation of the inside of the regulator is impossible in the case of strictly explosion-proof construction. Testing and inspection of the diaphragm condition must be performed by dismantling the regulator or by using expensive remote sensing systems operated by frontline workers. Therefore, we developed an inexpensive and easy to maintain system based on the indirect measurement of diaphragm vibration. However, a number of problems remain to be solved before its realization. The measurement system has an unacceptably low sensitivity to diaphragm vibration, which is not sufficient for easy fault diagnosis, because of the effects of nonlinearity of the regulator mechanism and strong noise intensity due to influence of the turbulent flow on the observed data. In addition, the change in the gas flow influences the observed data to a greater degree than the change in the vibration property caused by diaphragm deterioration. Thus, the amplitude of the obtained signals provides little information for correct classification. As a result, conventional methods or linear classification can not achieve high-performance diagnosis. Therefore, we need a more intelligent method.

The SVM has been applied to fault diagnosis for machines that have a vibrating or rotating structure [6]. However, a first-order polynomial kernel was used as a kernel function in the SVM, which is virtually the same as linear classification. The adoption of such methods is not efficient for the present fault detection problem because the difference in the observed signals between the normal and deteriorated regulator as measured by the newly developed system are insignificant. Moreover, the characteristic parameters obtained for individual regulators have large variance due to noise or nonlinearity. The present problem requires the application of a more powerful kernel in order to obtain high classification performance. Here, we focus on the Kullback-Leibler (KL) divergence, which is a criterion of the distance between two probability distributions. The KL divergence is often used to measure the distance from the true probability distribution to an arbitrary distribution in the framework of probability or information theory. The KL kernel method, in which the KL divergence is introduced as a kernel, was proposed for speech and image classification [7, 8]. This kernel is expected to become a good similarity measure for not only probabilistic distribution but also the normalized frequency spectrum.

In the present paper, we describe early fault diagnosis of the LP gas pressure regulator using the SVM. The vibrations of rubber diaphragms are measured as time series data using the newly developed system, and the frequency spectrums of the vibration are examined for further investigations. The diagnosis is carried out by frequency spectrum classification with training data sets of normal and deteriorated regulators. After explaining the LP gas regulator and the mechanism of diaphragm vibration, we consider the properties of the observation data and the method of preprocessing. Next, three polynomial kernels, Gaussian kernel, χ^2 kernel and KL kernel, are used to examine the SVM, and the conventional method of classification performance is shown. Figure 1 outlines the diagnosis methods considered in the present paper.

2. LP GAS PRESSURE REGULATOR AND THE NEWLY DEVELOPED MEASUREMENT SYSTEM

2.1. LP gas pressure regulator

Figure 2 shows the LP gas single-stage pressure regulator. Although there are various types of LP gas regulators, the fundamental structure of the pressure regulator is the same. We limit the present discussion to domestic LP gas single-stage pressure regulators, which are used to depressurize

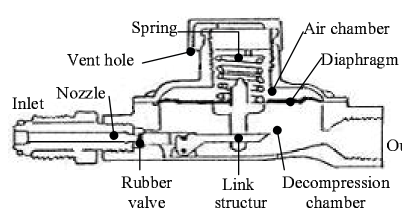

Fig. 2. LP gas single-stage pressure regulator

high-pressure gas in containers to low pressure for use in domestic gas appliances such as stoves. For this purpose, the regulator controls the divergence of the rubber valve and the nozzle to equalize the pressure in a decompression chamber and the pressure of an air chamber to a desirable output pressure. Regulators have such a feedback system.

2.2. Cause of pressure regulator failure

The causes and modes of failures were varied, for example, frozen water due to internal condensation in a cold region or reliquefaction in the air chamber. These phenomena cause disturbances in the feedback system of the regulator. However, these failures are detected early because the sound of abnormal diaphragm vibration can be detected by the human ear.

The damping coefficient of rubber, which is a characteristic of vibrational absorption, decreases with deterioration [9]. As the damping coefficient decreases, the vibration becomes difficult to regulate. The sound of the vibration can not be detected by the human ear because the changes in the regulator diaphragm vibration are subtle. In addition, the deterioration progresses gradually. Such deterioration of the rubber diaphragm is defined as an early fault in the preset paper.

2.3. Newly developed diaphragm vibration measurement system

The exterior of the regulator is sturdy in order to ensure explosion-proof construction, and it is difficult to measure the vibration of the diaphragm directly. This vibration can be measured using a remote measurement system. However, this is not practical because it requires expensive equipment and a highly skilled operator. We developed the measurement system illustrated in Figure 3. The mechanism of the proposed system is as follows. A microphone fixed to the upper wall inside a semi-closed vessel is used to detect changes in pressure based on the air flow generated due to the movement of the diaphragm. The microphone is inexpensive and small and has been used in a number of applications. Therefore, the proposed system is inexpensive and lightweight, allowing ease of use and portability.

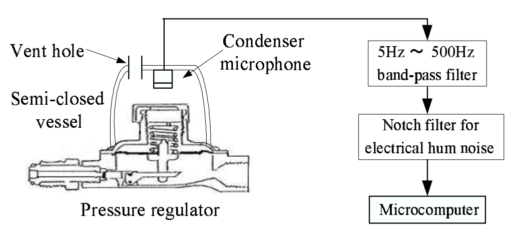

Fig. 3. Proposed system for early fault diagnosis of the pressure regulator

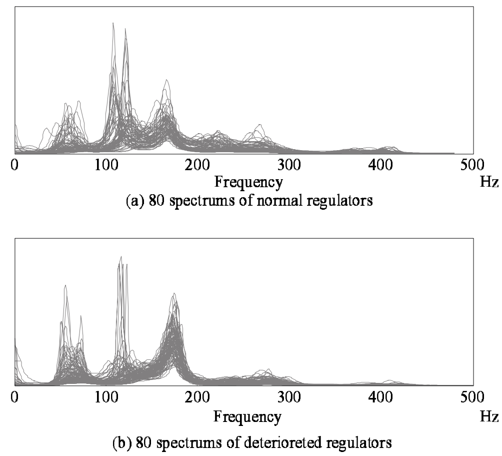

Fig. 4. Normalized spectrums of the diaphragm vibration of normal and deteriorated regulators

3. PROPERTIES OF OBSERVED DATA AND PREPROCESSING

In the measurement experiments, we examined four unused regulators (normal regulators), three of which had been expired for more than three years and one of which had been made to deteriorated artificially (deteriorated regulator). Measurement was performed four times for each the eight regulators for flow rates of 200 L/h, 400 L/h, 600 L/h, 800 L/h and 1,000 L/h with a sampling time of 1 millisecond. We obtained a total of 160 data, including 80 normal regulator data and 80 deteriorated regulator data.

Change in the gas flow strongly influence the observed data in the proposed measurement system, as mentioned above. In order to suppress the unfavorable effect on classification caused by amplitude differences of the gas flow, we use normalized spectrums as a feature for classification. Figure 4 shows the normalized spectrums for the 80 normal regulator data and the 80 deteriorated regulator data, respectively. These spectrums were estimated by the AR method with an optimal number of coefficients obtained by the AIC. Figure 4 shows that it is not easy to determine which peak has a property associated with the change in the diaphragm condition because all of the spectrums have multiple peaks. For this reason, the application of the half-power method to these spectrum analysis seems impractical. Moreover, the spectrums have an insignificant difference between the normal and deteriorated regulators. Therefore, we have to adopt a more intelligent method in order to obtain a high classification performance.

4. SUPPORT VECTOR MACHINE AND KULLBACK-LEIBLER KERNEL

4.1. Support Vector Machine

In this subsection, we provide a brief review of the SVM. The SVM is a linear learning machine for binary classification. The infinite nuber of discriminant hyperplane for classification generally exists if the given data set is linearly separable. The advantages of the SVM are that one discriminant hyperplane can be created that has a maximum margin (the distance between the discriminant hyperplane and the nearest point in training data to the hyperplane) with the given training data for a classifier and that the discriminant hyperplane is specified uniquely. Furthermore, using the kernel method, the SVM becomes a powerful nonlinear classifier.

Here, consider a binary classification that has a training set consisting of N pairs of inputs and outputs $(\boldsymbol{x}_i, y_i) \in X \times Y, Y = \{-1, 1\} (i = 1, \cdots, N)$. A discriminant hyperplane with the weighting coefficient $\boldsymbol{w}$, given by

$$f(\boldsymbol{x}) =< \boldsymbol{w} \cdot \boldsymbol{x} > +\beta \tag{1}$$

is determined by solving a constrained optimization problem in dual space

$$\begin{aligned} &maximize \quad \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} y_i y_j \alpha_i \alpha_j < \boldsymbol{x}_i \cdot \boldsymbol{x}_j > \\ &subject\ \ to \quad \sum_{i=1}^{N} y_i \alpha_i = 0, 0 \le \alpha_i \le C \end{aligned} \tag{2}$$

where a_i is the Lagrange multiplier, C is a regularization parameter that controls the trade-off between the complexity of the discriminant hyperplane and the errors of the SVM on training, and $< \cdot >$ is the inner product. This is referred to as the soft margin SVM. For expansion of the SVM to nonlinear classifiers, nonlinear mapping from the original input space X to the feature space Z $\Phi : X \rightarrow Z$ is introduced, and the similarity of $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is evaluated by the inner product $< \Phi(\boldsymbol{x}_i) \cdot \Phi(\boldsymbol{x}_j) >$. Here, the function

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) =< \Phi(\boldsymbol{x}_i) \cdot \Phi(\boldsymbol{x}_j) > \tag{3}$$

is referred to as the kernel function. For easy calculation of the inner product and in order to prevent a substantial increase in computational complexity, the kernel function is usually chosen as a function that can be computed without explicit calculation of the inner product in the feature space. The SVM becomes a nonlinear classifier by introducing a kernel function to Equation (2), as follows:

$$< \boldsymbol{x}_i, \boldsymbol{x}_j > \rightarrow K(\boldsymbol{x}_i, \boldsymbol{x}_j).$$

That is to say, introducing the kernel function is equivalent to linear classification in a high-dimensional feature space

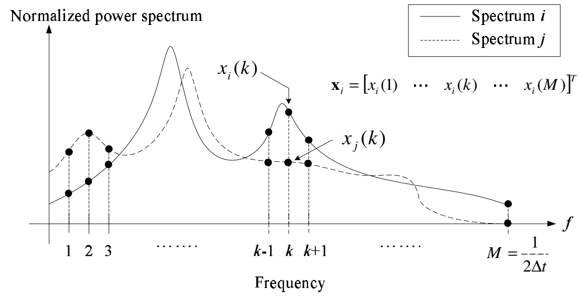

Fig. 5. Discretized frequency spectrum for constitution of the feature vector

mapped by a certain nonlinear function. Therefore, classification performance of the nonlinear SVM heavily depends on the choice of the kernel function.

4.2. Kullback-Leibler kernel

The KL divergence gives a metric between two probability distributions. The continuous form of the KL divergence between $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$ with random variable $\boldsymbol{x}$ is defined as

$$D(p(\boldsymbol{x}), q(\boldsymbol{x})) = \int_{-\infty}^{\infty} p(\boldsymbol{x}) \log \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})} d\boldsymbol{x}. \tag{4}$$

The kernel function must be a symmetric function [1, 2]. However, the KL divergence is not symmetric, and the symmetric KL divergence is considered to be

$$\begin{aligned} SD(p(\boldsymbol{x}), q(\boldsymbol{x})) \quad = \quad & \int_{-\infty}^{\infty} p(\boldsymbol{x}) \log \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})} d\boldsymbol{x} \\ + \quad & \int_{-\infty}^{\infty} q(\boldsymbol{x}) \log \frac{q(\boldsymbol{x})}{p(\boldsymbol{x})} d\boldsymbol{x} \end{aligned} \tag{5}$$

and the KL kernel by [7] is defined as

$$K_{KL}(p(\boldsymbol{x}), q(\boldsymbol{x})) = e^{-a\{SD(p(\boldsymbol{x}), q(\boldsymbol{x}))+b\}} \tag{6}$$

where a and b are constants. This kernel has a property whereby the generative probability and discriminate model can be combined.

4.3. Modification of the KL kernel

The KL kernel is usually recognized as a kernel that measures the similarity of two probability distributions. In speech and image classification applications [7, 8], the KL kernel measures the similarity of two Gaussian Mixture Models fitted to the decomposed frequency components from speech or image data by Discrete Cosine Transform (DCT). However, in the present study, we do not consider the probability density because we deal not with the probability function but with the frequency spectrum. Figure 5 shows a schematic illustration of the components of the feature vector, where Δt is the sampling time and M is the number of elements of the feature vector. The feature vector consists of the discretized frequency spectrum, as $\boldsymbol{x}_i = [x_i(1) \cdots x_i(k) \cdots x_i(M)]^T$. In the present applications, the KL kernel is defined as

$$K_{KL}(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-a\{SD(\boldsymbol{x}_i, \boldsymbol{x}_j)+b\}} \tag{7}$$

$$SD(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sum_{k=1}^{M} \left\{ x_i(k) \log \frac{x_i(k)}{x_j(k)} + x_j(k) \log \frac{x_j(k)}{x_i(k)} \right\}. \tag{8}$$

5. EXPERIMENTS

The experiments were conducted using 80 data of normal regulators and 80 data of deteriorated regulators, as explained in Section 3. A total of 1024 data points (approximately one seconds of time series data) were selected from the data for the feature vector $\boldsymbol{x}_i$, and these data were used for classification.

For comparison of the KL kernel with some popular kernels and the conventional diagnosis method, we examined the classification performance of these methods using six types of features. The first feature that we selected is the AR coefficient. The AR coefficient from the first order to the 30th order as estimated by the Yule-Walker method was selected as the feature vector $\boldsymbol{x}_i = [x_i(1), \cdots, x_i(30)]^T$. This feature is not a spectrum. The reason why we selected this feature is that the AR coefficient is often used as a feature for time series classification. The second feature that we selected is the periodogram. The normalized frequency components from 1 Hz to 500 Hz in 1-Hz increments estimated by DFT were selected as a feature $\boldsymbol{x}_i = [x_i(1), \cdots, x_i(500)]^T$. The Nyquist frequency of the set sampling time is 500 Hz. The third feature that we selected is the smoothed spectrum obtained by the AR method. The normalized spectrum estimated by the optimal order AR method in terms of the AIC was selected. This feature consists of the components from 1 Hz to 500 Hz, as in the case of the periodogram. The remaining features that we selected are smoothed periodograms obtained by 5-point, 30-point and 60-point moving averages. The elements of these feature vector are the same as in the case of the periodogram. The kernel functions used with the SVM for these six input features are the first-order polynomial kernel (linear classification), the second-order polynomial kernel, the third-order polynomial kernel, the Gaussian kernel, the χ^2 kernel, and the KL kernel. In addition, we examined the performance of the half-power method, which is a conventional diagnosis method.

Table 1. Accuracy percentage of the combination of the classifier and the features

	ARc	PG	AR	MA1	MA2	MA3
P1	76.65	69.35	81.22	76.28	79.15	77.23
P2	83.75	69.97	81.91	78.24	80.01	78.36
P3	84.52	70.31	83.94	77.91	80.48	79.30
GA	76.85	68.89	82.90	77.14	80.27	79.05
χ^2	76.04	69.13	89.77	76.69	80.44	79.31
KL	×	80.61	**92.41**	83.44	89.16	82.79
HP	×	×	49.62	54.26	46.53	56.33

Table 2. Accuracy percentage of the fitted Gaussian mixture and the spectrum for the KL kernel

GM3	GM4	GM5	AR
88.34	90.02	88.95	**92.41**

A total of 40 data extracted for the normal regulator and 40 data extracted for the deteriorated regulator were randomly selected from each data set and assigned as a training data set. The remaining 80 data were then used for validation. We repeated this operation 100 times. The SVM and kernel parameters were fixed such that the accuracy rates are as good as possible in these experiments. Table 1 shows the average accuracy percentage in the 100 operations. In Table 1, ARc, PG, AR, MA1, MA2 and MA3 indicate the AR coefficient, the normalized periodogram, the normalized spectrum estimated by the AR method and the smoothed normalized-spectrum estimated by the 5-point, 30-point and 60-point moving average methods, respectively. In addition, P1, P2, P3, GA, χ^2, and KL denote the first-order polynomial, the second-order polynomial, the third-order polynomial, the Gaussian, the χ^2, and the KL kernel, respectively. Also, HP denotes the use of the half-power method, and "×" indicates that a calculation is impossible for this combination of feature and kernel. The combination of the KL kernel with an input feature of the normalized spectrum by the AR method provides the best performance.

The comparison of the probability distribution function fitted Gaussian mixture and the estimated spectrum to the KL kernel was also demonstrated. Table 2 shows the results for the correct answer percentage of the mixture of 3, 4 and 5 Gaussians of diagonal covariance and the estimated frequency spectrum. In Table 2, GM3, GM4 and GM5 denote Gaussian mixtures of 3, 4 and 5 normal distributions. These results also show that an input feature of the frequency spectrum attains the best performance.

6. DISCUSSION

The experimental results obtained herein imply three useful advantages. The first is the excellence of the KL kernel, as demonstrated in the first experiment. The KL kernel shows the best performance among all of the features considered herein (except for the case in which calculation is impossible). This result shows that the KL kernel has an advantage in similarity calculation, not only for probability distributions, but also for normalized frequency spectrums. The second advantage is that high classification performance is obtained with an input feature of a moderate smoothed spectrum, rather than a periodogram, as shown in the first experiment. Actually, spectrums estimated by the AR and 30-point moving average methods showed high accuracy rates. The input of the periodogram shows a poor performance compared to the other inputs. The reason for this is thought to be the high sensitivity of the spectrum estimated by DFT with respect to the noise effect. Therefore, suitable input data representation or estimation is needed for correct classification. As in the case of the periodogram, the 5-point moving average lacks smoothness and the 60-point moving average is too smooth. An optimal number of data points for the moving average procedure must be chosen by the cross-validation technique, for example. The AR method is convenient for practical use because it can be used to automatically determine the optimal order in terms of the AIC. The third advantage, as shown by the second experiment, is better performance for the frequency spectrum than for Gaussian mixtures fitted as data. The KL kernel is generally expected to yield a good measure of similarity of spectrum if the input feature satisfies the following condition:

$$\sum_{k=1}^{M} \boldsymbol{x}_i(k) = \sum_{k=1}^{M} \boldsymbol{x}_j(k). \tag{9}$$

7. CONCLUSION

The present paper described early fault diagnosis of the LP gas pressure regulator using the SVM with the KL kernel. Suppliers of LP gas must be able to detect abnormal conditions of LP gas instruments, such as regulators, in order to ensure safety. We proposed a vibration measurement system for fault diagnosis of the internal rubber diaphragm of a regulator. However, the use of conventional diagnosis methods in the proposed system is inefficient due to problems such as strong noise and nonlinearity with respect to the observation method. In order to solve these problems, we applied the SVM with the KL kernel based on the KL divergence to spectrum classification. The KL kernel is used to measure the similarity between two probability distributions estimated from given data. The estimated normalized frequency spectrum was used as an input feature to the KL kernel as in the case of probability distribution. For the problem considered herein, the use of the KL kernel provided a high accuracy rate compared to the use of polynomial, Gaussian and χ^2 kernels, and conventional fault diagnosis methods by spectrum analysis. We also demonstrated that the proposed method, which employs the discretized frequency spectrum directly as a feature vector, is more efficient than the method that uses the KL kernel with the Gaussian Mixture Model. High classification performance was achieved, and a correct answer rate of more than 92% was attained using only an inexpensive sensor and the machine learning method. The proposed method can be applied to other spectrum classification problems without limiting the generality if the spectrums are normalized.

8. REFERENCES

[1] J. S-Taylor and N. Cristianini, *Kernel methods for pattern analysis*, Cambridge University Press, 2004.

[2] N. Cristianini and J. S-Taylor, *An introduction to Support Vector Machines and other kernel-based learning methods*, Cambridge University Press, 2004.

[3] M. Ge, R. Du, G. Zhang and Y. Xu, "Fault diagnosis using support vector machine with an application in sheet metal stamping operations," *Mechanical System and Signal Processing.*, vol. 18, pp. 143–159, 2004.

[4] C Saunders, A. Gammerman, H Brown and G Donald "Application of Support Vector Machines to Fault Diagnosis and Automated Repair," *Proc. Of the 11th Int. Workshop on Principles of Diagnosis.*, 2000.

[5] *Harris's Shock and Vibration Handbook 5th ed*, ed. C. M. Harris and A. G. Piersol McGraw-Hil, 2002.

[6] S. Poyhonen, M. Negrea, A. Arkkio, H. Hyotyniemi and H. Koivo, "Fault diagnostics of an electrical machine with multiple support vector classifiers," *Proc. the 2002 IEEE International Symposium Intelligent Control.*, pp. 373–378, 2002.

[7] P. J. Moreno, P. P. Ho and N. Vasconcelos, "A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications ," *In Advances in Neural Information Processing Systems.* 2003.

[8] A. B. Chan and N. Vasconcelos, "Probabilistic Kernels for the Classification of Auto-Regressive Visual Processes," *Proceedings of the 2005 IEEE Computer Society Conference CVPR.*, vol. 1, pp. 846–851, 2005.

[9] M. F. Ashby *Materials Selection in Mechanical Design*, Elsevier, 1999.