Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Sound and Vibration 287 (2005) 25–43
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

Cavitation detection of butterfly valve using support vector machines

Bo-Suk Yang[a,*], Won-Woo Hwang[a], Myung-Han Ko[b], Soo-Jong Lee[a]

[a]*Intelligent Mechanics Laboratory, School of Mechanical Engineering, Pukyong National University, San 100, Yongdang-dong, Nam-gu, Busan 608-739, South Korea*
[b]*Paldang Regional Office, Korea Water Resources Corporation, San 13, Bealmi-dong, Hanam, Kyonggi 465-130, South Korea*

Received 23 June 2003; received in revised form 27 April 2004; accepted 25 October 2004
Available online 23 December 2004

Abstract

Butterfly valves are popularly used in service in the industrial and water works pipeline systems with large diameter because of its lightweight, simple structure and the rapidity of its manipulation. Sometimes cavitation can occur, resulting in noise, vibration and rapid deterioration of the valve trim, and do not allow further operation. Thus, monitoring of cavitation is of economic interest and is very important in industry.

This paper proposes a condition monitoring scheme using statistical feature evaluation and support vector machine (SVM) to detect the cavitation conditions of butterfly valve which used as a flow control valve at the pumping stations. The stationary features of vibration signals are extracted from statistical moments. The SVMs are trained, and then classify normal and cavitation conditions of control valves. The SVMs with the reorganized feature vectors can distinguish the class of the untrained and untested data. The classification validity of this method is examined by various signals acquired from butterfly valves in the pumping stations. And the classification success rate is compared with that of self-organizing feature map neural network (SOFM).

*Corresponding author. Tel.: +82 51 620 1604; fax: +82 51 620 1405.
E-mail address: bsyang@pknu.ac.kr (B.-S. Yang).

0022-460X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsv.2004.10.033

1. Introduction

Most pumping stations have evolved to calendar scheduling of maintenance: valves are maintained on a cyclic basis even if no signs of trouble are detected. However, cyclic scheduling may leave out components that will fail before their planned service date if signs of deterioration are not clearly visible at an earlier inspection. A less expensive and more effective valve maintenance schedule can be implemented only if the selection of valves to be serviced is made on a justifiable priority-of-need basis. The purpose of the valve monitoring is to provide the means to do that.

Industrial piping systems can experience severe vibration and noise caused by internal flow of the conveyed fluid. Valves can result in annoying and even unbearable noise and high vibration [1]. Engineers typically lavish much attention on pumps and little on valves which is just as important for the proper functioning of a pumping station. Flow control valves which are the most important valves in a pumping station, are used to modulate flow by operating in a partly open position, thus creating a high head-loss or pressure differential between upstream and downstream locations. Such operation may create cavitation and noise.

Usually, butterfly, ball, cone or plug valves are used for flow control valves, where energy costs are important. Butterfly valves are popularly used in service in the industrial and water works pipeline systems with large diameter because of its lightweight, simple structure and the rapidity of its manipulation. Butterfly valves can be designed for leak-proof shut-off, but leakage is significant without a resilient seat [2].

When the butterfly valve is used as a flow control valve, cavitation phenomena sometimes occur in the range of higher flow rate because of the small valve opening. When a liquid passes through a pipe, the velocity is comparatively low because of the relatively large cross section. As the liquid passes through a valve seat, its velocity increases. An increase in velocity increases dynamic pressure, which reduces the static pressure. If the velocity is high enough, the pressure at the valve seat can drop below the vapor pressure of the liquid and form vapor bubbles. The downstream static pressure is normally higher than the vapor pressure of the liquid. Therefore, the bubbles or cavities of vapor implode. When a bubble implodes, all the energy is concentrated into a very small area. This creates tremendous pressure of thousands of psi in the small area, generating minute shock waves. These shock waves pound on the solid portions of the valve. Repeated implosions on a small surface eventually cause fatigue of the metal and wear away this surface. Fig. 1 shows an example of the pitting corrosion at the rear surface of valve body and valve seat of nozzle side caused by cavitation at valve opening of 20%. Cavitation is a potential danger, especially when valves operate at low opening, and may damage the valve very rapidly. It is suggested that the pitting and erosion is accelerated by simultaneous chemical attack or that the high impact pressure causes locally high temperatures that accelerate pitting.

Sometimes cavitation can occur, resulting in noise, vibration and rapid deterioration of the valve trim, and do not allow further operation. Thus, the monitoring of cavitation is of economic interest. Usually the valve trim is destroyed prematurely, but the valve body and the pipe also can be affected as far away as 20 diameters downstream of the valve [3].

In the case of two-class classification problems, the main objective is to find a, in general, nonlinear optimal separating surface between the two classes (cavitation and normal conditions), starting from a collection (training set) of examples of signals belonging to the two classes. In this

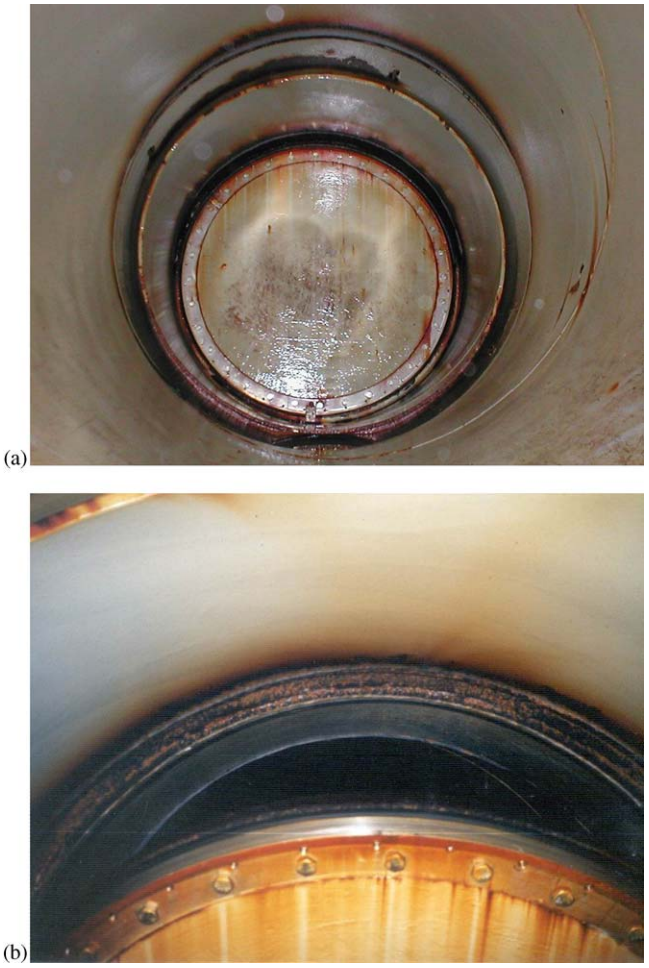

Fig. 1. Occurrence of pitting corrosion due to cavitation. (a) Rear surface of valve body; (b) valve seat of nozzle side.

paper, we adopt a new category of universal feed-forward network known as support vector machines (SVMs) introduced by Vapnik [4]. SVMs, based on statistical learning theory, are a comparatively recent development, although their origins can be tracked back to the late 1960s [5,6]. SVMs have been gaining acceptance in the machine learning, computer vision and pattern recognition communities for their high accuracy and good generalization capability [7]. The main difference between artificial neural networks (ANNs) and SVMs is in the principle of risk minimization [8]. In case of SVMs, structural risk minimization principle is used minimizing an upper bound on the expected risk whereas in ANNs, traditional empirical risk minimization is used minimizing the error on the training data. The difference in risk minimization leads to better generalization performance for SVMs than ANNs. The possibilities of using SVMs in machine condition monitoring application are being considered only recently by Jack and Nandi [9] and Samanta [10]. However, there are still relatively few 'real' engineering applications based on them, and only a few in the field of process/ condition monitoring [11].

The goal of this work is to present a cavitation detection scheme based upon SVM for the butterfly valves in large pumping stations. This paper provides a comparison between the two classification algorithms, SVMs and ANNs such as self-organizing feature map (SOFM) networks, and shows results through utilizing the vibration signals of the butterfly valves in pumping stations with different valve opening conditions. A statistical method is used to extract features from the measured signals. The real data is used to demonstrate the capability of this system in valve cavitation detection. The results show the effectiveness of the extracted features from the acquired signals in classification of the valve condition.

2. Test facility

The test valves are the inflow valves for gauging well which used in wide area water works pipeline networks. These valves are installed horizontally at circular pipe with a diameter of 1000 mm and located at 2 km away from upstream pressurization station. The pipe pressure at valve upstream is 0.245 MPa for valve shutdown condition. Table 1 shows the flow rate and pressure according to the valve opening condition. The schematic diagram of a butterfly valve under investigation is shown in Fig. 2. The valve body is stroked by an electric motor-driven diaphragm type actuator. A butterfly valve is a quarter-turn valve in which a disc is rotated on a shaft so that the disc seats on a ring in the valve body. Usually the valve seat is an elastomer bonded or fastened either to the vane or to the body. The vane protrudes into the adjacent piping when in the open position. The valve can be actuated with a simple lever attached to the plug shaft. In service, the control range for the vane angle is about 20–60°.

3. Data acquisition and feature extraction

3.1. Data acquisition

Two accelerometers were directly mounted vertically and horizontally at the valve body to measure the vibration signals at various valve opening from 20% to 60%. Seventy continuous measurements were recorded for 5 min for each condition. The maximum acquisition frequency was 10 kHz and the sampling number was 16384. A mobile DSP analyzer was used to perform the data acquisition and the data were stored in a notebook computer. Fig. 3 shows frequency

Table 1
Flow rate and pressure according to valve opening

Valve opening (%)	Flow rate (m^3/min)	Pressure (MPa)
10	642	0.240
20	2981	0.229
30	4204	0.178
60	6542	0.084

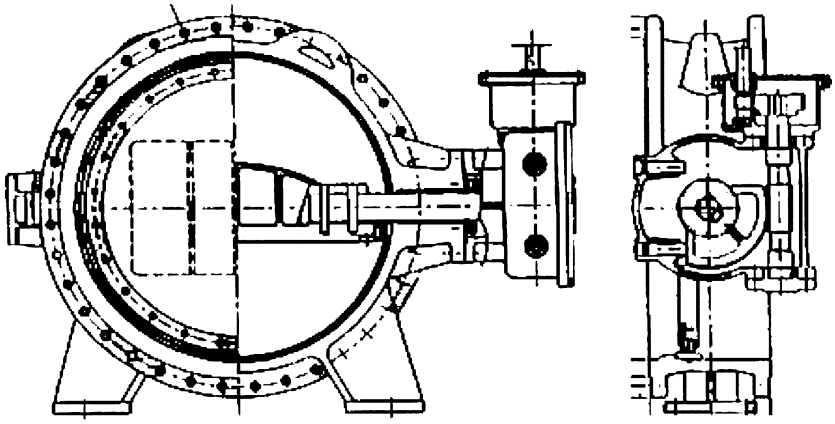

Fig. 2. Cross-sectional view of a butterfly valve.

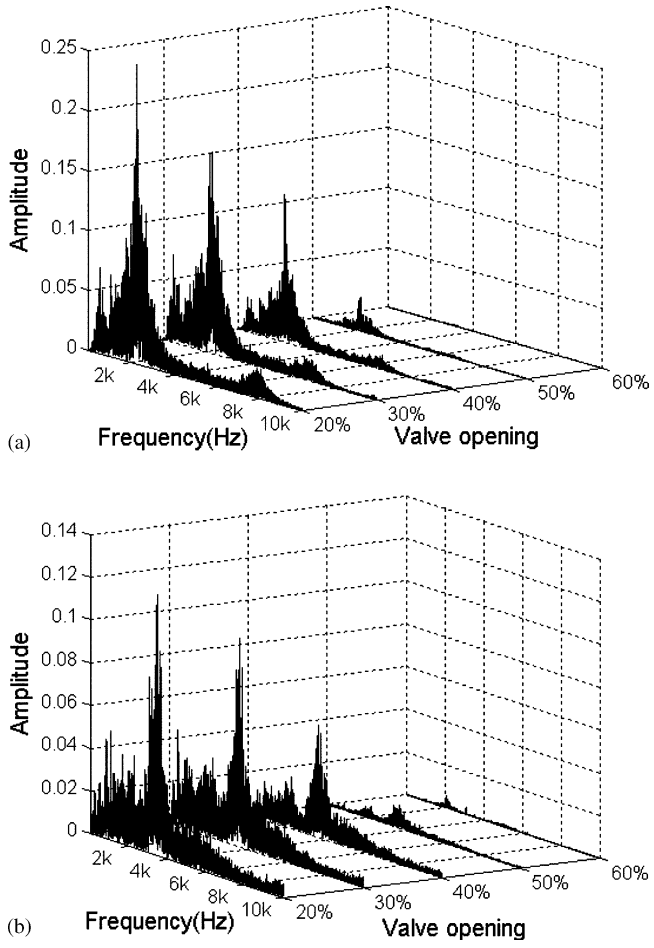

Fig. 3. Frequency spectrum according to valve opening. (a) Vertical direction; (b) horizontal direction.

spectrums of the raw vibration signals collected from the valve body for different valve opening. In Fig. 3, one can observe the emergence of peaks of high-frequency structural resonance. The center of frequency distribution is almost constant about 2.2 kHz vertically and 3.5 kHz horizontally from normal state to small opening conditions. When the cavitation occurs, the fluctuating pressure is directly applied to the valve body due to sudden implode of the bubble. As the liquid passes through in vertical direction, the amplitude of vertical direction is larger than that of horizontal direction. The magnitude of frequency spectrum tends to increase dramatically when the cavitation occurs due to the decrease of the valve opening from 60% to 20%. However, it is difficult to differentiate the vibration energy change from spectrum. The energy spectrum is distributed in a wide range from 100 Hz to 8 kHz for the vibration signals collected from the valve body. From the time series waveform, no conspicuous difference exists among the different valve openings. There is a need to come up with a feature extraction method to classify them.

3.2. Statistical feature extraction

Although the time series data contain abundant feature information, the important part cannot be shown intuitively, and much unnecessary information is also contained. Therefore, the feature extraction is essential for effective estimation of valve conditions. Several statistical parameters, calculated in the time domain, are generally used to define average properties of acquired data. Many of these features are based on moments and therefore the method of estimating these and their relationship with the distribution of the random variable. In most cases, the probability density function (pdf) can be decomposed into its constituent moments. If a change in condition causes a change in the pdf of the signal then the moments may also change therefore monitoring these can provide diagnostic information.

In this paper, the four basic parameters, mean value μ, standard deviation σ, kurtosis k and shape factor s, are used as the statistical feature. For a given data set x_i, $i = 1, \ldots, N$, these are defined as follows:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{1}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}, \tag{2}$$

$$k = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^4}{\sigma^4}, \tag{3}$$

$$s = \frac{\sigma}{\mu}, \tag{4}$$

where N is the number of the data points.

The kurtosis gives an indication of the proportion of samples that deviate from the mean by a small value compared to those which deviate by a large value. The kurtosis is determined not by the magnitude of the waveform but by the shape of the waveform. It can be either positive or

negative, and is very close to unity for a normal distribution. Random variables that have a negative kurtosis are called subgaussian, and those with positive kurtosis are called supergaussian. Supergaussian variables have typically a spiky pdf with heavy tails, i.e. the pdf is relatively large at zero and large values of the variable, while being small for intermediate values. On the other hand, subgaussian variables have typically flat pdf, which is rather constant near zero, and very small for larger values of the variable. The zero mean Gaussian distributed variable has a kurtosis of 3. These statistical parameters may be used to perform a quick check of the changes in the statistical behavior of a signal [12]. Figs. 4 and 5 show plots of some of these features extracted from the vibration acceleration signals. At the valve opening of 60% where values of mean, root mean square (rms) and shape factor are very low, kurtosis has large deviation. Cavitation does not occur in this condition. For the valve opening of 50%, these values rise up and kurtosis approaches to 3. This means the cavitation inception occurs. For 40% and less, shape factor is

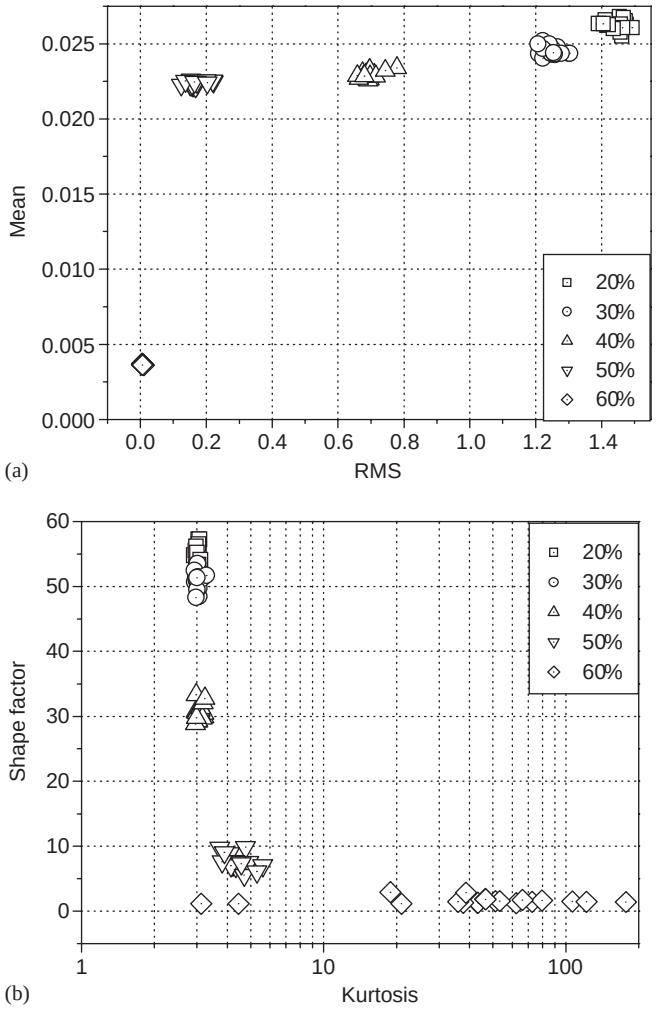

Fig. 4. Feature characteristics according to valve opening for vertical direction.

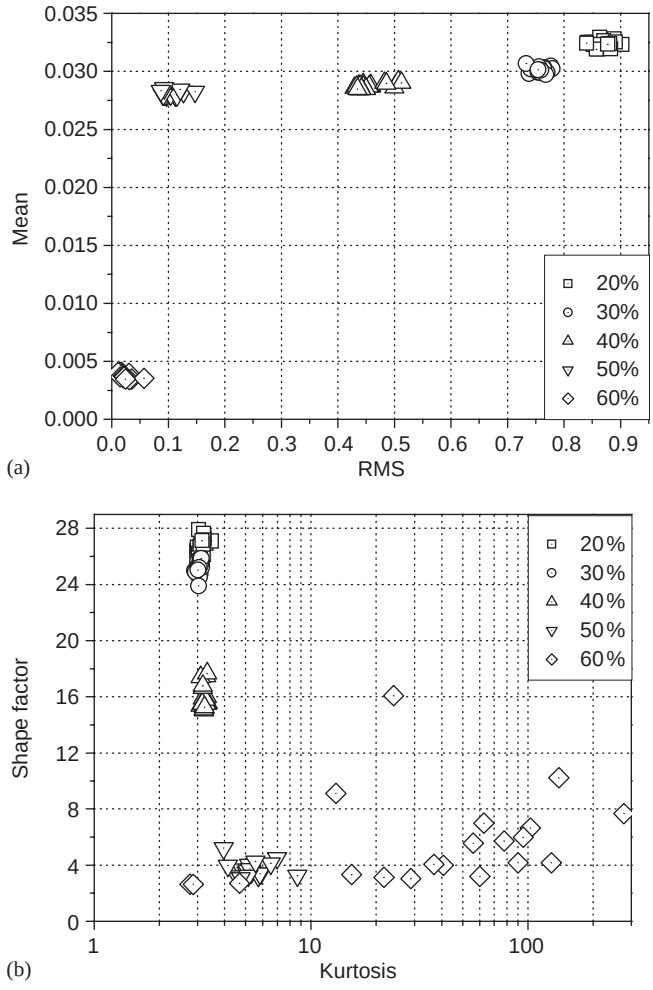

Fig. 5. Feature characteristics according to valve opening for horizontal direction.

rapidly increased and kurtosis is a constant of 3, cavitation appears at the whole valve edge, that is, supercavitation begins from valve opening of 40% [13].

4. Support vector machines (SVMs)

For detailed tutorials on the subject the reader can refer to Refs. [4,6,7,14] and references cited therein. In this section a brief outline of the method will be described. The SVM attempts to create a line or hyperplane between two sets of data for classification. In a 2D situation, the action of the SVM can be explained easily without any loss of generality. Fig. 6 shows how to classify a series of points into two different classes of data, class A (circles) and class B (squares). The SVM attempts to place a linear boundary represented by a solid line between the two different classes and orients

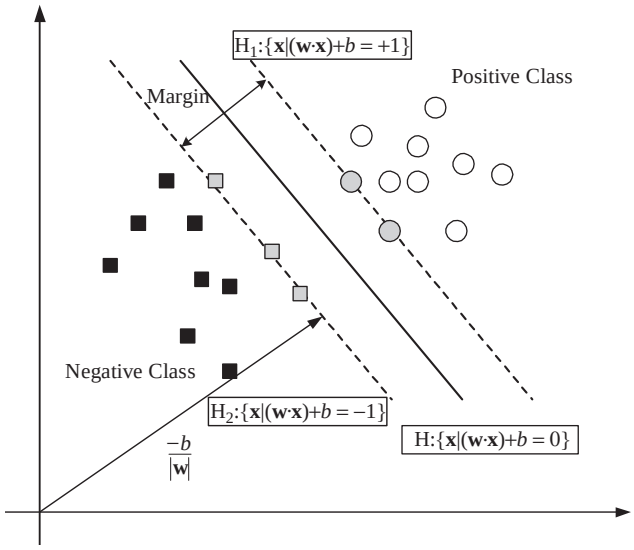

Fig. 6. An example of classification of two classes by SVM.

it in such a way that the margin represented by dotted lines is maximized. The SVM tries to orient the boundary such that the distance between the boundary and the nearest data point in each class is maximal. The boundary is then placed in the middle of this margin between the two points. The nearest data points are used to define the margins and are known as support vectors (SVs) represented by gray circle and square. Once the SVs are selected, the rest of the feature set can be discarded, since the SVs have all the necessary information for the classifier [10].

Let (x_i, y_i), with $i = 1, \ldots, N$; be a training example set S; each example $x_i \in R^N$ belongs to a class by $y_i \in \{-1, 1\}$. The goal is to define a hyperplane which divides S, such that all the points with the same label are on the same side of the hyperplane while maximizing the distance between the two classes A, B and the hyperplane. The boundary can be expressed as follows:

$$\mathbf{w} \cdot \mathbf{x} + b = 0, \quad \mathbf{w} \in R^N, \quad b \in R, \tag{5}$$

where the vector $\mathbf{w}$ defines the boundary, $\mathbf{x}$ is the input vector of dimension N and b is a scalar threshold. At the margins, where the SVs are located, the equations for classes A and B, respectively, are as follows:

$$\mathbf{w} \cdot \mathbf{x} + b = 1, \quad \mathbf{w} \cdot \mathbf{x} + b = -1. \tag{6}$$

As SVs correspond to the extremities of the data for a given class, the following decision function can be used to classify any data point in either class A or B:

$$f(\boldsymbol{x}) = \text{sign}(\mathbf{w} \cdot \mathbf{x} + b). \tag{7}$$

The optimal hyperplane separating the data can be obtained as a solution to the following optimization problem [15]:

Minimize

$$\tau(\mathbf{w}) = 1/2\|\mathbf{w}\|^2 \tag{8}$$

subject to

$$y_i(\mathbf{w}\cdot\mathbf{x}_i+b)\geqslant 1,\quad i=1,2,\ldots,N, \tag{9}$$

where N is the number of training sets.

However, if the only possibility to access the feature space is via dot products computed by the kernel, we cannot solve Eq. (8) directly since $\mathbf{w}$ lies in that feature space. But it turns out that we can get rid of the explicit usage of $\mathbf{w}$ by forming the dual optimization problem [16]. Introducing Lagrange multipliers $\alpha_i\geqslant 0$, $i=1,2,\ldots,N$, one for each of the constraints in Eq. (9), we get the following Lagrangian:

$$L(\mathbf{w},b,\boldsymbol{\alpha})=\frac{1}{2}\|\mathbf{w}\|^2-\sum_{i=1}^{N}\alpha_i y_i(\mathbf{w}\cdot\mathbf{x}_i-b)+\sum_{i=1}^{N}\alpha_i. \tag{10}$$

The task is to minimize Eq. (10) with respect to $\mathbf{w}$ and b, and to maximize it with respect to α_i. At the optimal point, we have the following saddle point equations:

$$\frac{\partial L}{\partial \mathbf{w}}=0,\quad \frac{\partial L}{\partial b}=0, \tag{11}$$

which translate into

$$\mathbf{w}=\sum_{i=1}^{N}\alpha_i y_i\mathbf{x}_i,\quad \sum_{i=1}^{N}\alpha_i y_i=0. \tag{12}$$

From the first equation of Eq. (12), we find that $\mathbf{w}$ is contained in the subspace spanned by the $\mathbf{x}_i$. By substituting Eq. (12) into Eq. (10), we get the dual quadratic optimization problem:

Maximize

$$L_D(\alpha)=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i,j}^{N}\alpha_i\alpha_j y_i y_j\mathbf{x}_i\cdot\mathbf{x}_j \tag{13}$$

Subject to

$$\alpha_i\geqslant 0,\quad i=1,2,\ldots,N, \tag{14}$$

$$\sum_{i=1}^{N}\alpha_i y_i=0. \tag{15}$$

Thus, by solving the dual optimization problem, one obtains the coefficients α_i which is required to express the $\mathbf{w}$ to solves Eq. (8). This leads to the nonlinear decision function

$$f(\mathbf{x})=\operatorname{sign}\left(\sum_{i=1}^{N}\alpha_i y_i(\mathbf{x}_i\cdot\mathbf{x})+b\right). \tag{16}$$

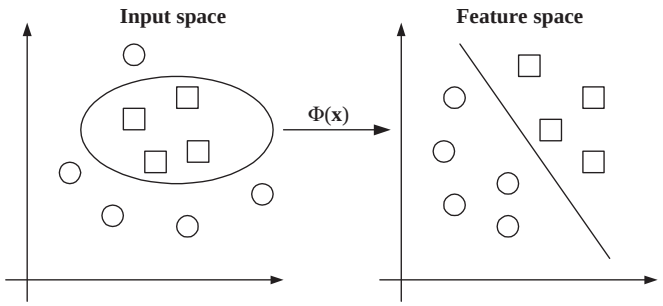

Fig. 7. Transformation to linear feature space from nonlinear input space.

In cases where the linear boundary in the input spaces are not enough to separate the two classes properly, it is possible to create a hyperplane that allows a linear separation in the higher dimension. In SVMs, this is achieved through the use of a transformation $\Phi(\mathbf{x})$ that converts the data from an N-dimensional input space to Q-dimensional feature space:

$$\mathbf{s} = \Phi(\mathbf{x}), \tag{17}$$

where $\mathbf{x} \in R^N$ and $\mathbf{s} \in R^Q$.

Fig. 7 shows the transformation from input space to feature space where the nonlinear boundary has been transformed into a linear boundary in feature space. Substituting the transformation Eq. (17) in Eq. (7) gives the decision function as

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}_i)) + b\right). \tag{18}$$

A kernel function $K(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x}) \cdot \Phi(\mathbf{y})$ is used to perform the transformation into higher-dimensional feature space. The basic form of SVM is obtained after substituting the kernel function in the decision function equation (18) as follows:

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b\right). \tag{19}$$

Any function that satisfies Mercer's theorem [4] can be used as a kernel function to compute a dot product in feature space. There are different kernel functions used in SVMs, such as linear, polynomial, Gaussian RBF. The selection of an appropriate kernel function is important, since the kernel function defines the feature space in which the training set examples will be classified. The definition of legitimate kernel function is given by Mercer's theorem: The function must be continuous and positive definite. In this work, linear, polynomial, χ^2 and Gaussian RBF kernel functions were evaluated and formulated as shown in Table 2.

In Table 2, d is the degree of the polynomial. σ denotes the width of the RBF kernel parameter and can be determined in general by an iterative process selecting an optimum value based on the full feature set [9]. This kernel is also well accepted for constructing SVMs and provides excellent results for real-world applications [17].

Table 2
Formulation for used kernel functions

Kernel	$K(\mathbf{x}, \mathbf{y})$
Linear	$\mathbf{x} \cdot \mathbf{y}$
Polynomial	$(\mathbf{x} \cdot \mathbf{y} + 1)^d$
Gaussian RBF	$\exp\{-(\|\mathbf{x} - \mathbf{y}\|^2/2\sigma^2)\}$
χ^2	$(\mathbf{x} - \mathbf{y})^2/(\mathbf{x} + \mathbf{y})$

5. Sequential minimal optimization (SMO) algorithm

To solve the SVM problem one has to solve the quadratic programming (QP) problem of Eq. (13) under the constraints Eqs. (14) and (15). Vapnik [18] describes a method which used the projected conjugate gradient algorithm to solve the SVM-QP problem, which has been known as "chunking". The chunking algorithm uses the fact that the value of the quadratic form is the same if you remove the rows and columns of the matrix that corresponds to zero Lagrange multipliers. Therefore, chunking seriously reduces the size of the matrix from the number of training examples squared to approximately the number of non-zero Lagrange multipliers squared. However, chunking still cannot handle large-scale training problems, since even this reduced matrix cannot fit into memory [19]. Osuna et al. [20] proved a theorem which suggests a whole new set of QP algorithms for SVMs. The theorem proves that the large QP problem can be broken down into a series of smaller QP sub-problems. Sequential minimal optimization (SMO) proposed by Platt [21] is a simple algorithm that can be used to solve the SVM-QP problem without any additional matrix storage and without using the numerical QP optimization steps. This method decomposes the overall QP problem into QP sub-problems using the Osuna's theorem to ensure convergence. In this paper the SMO is used as a solver and detail descriptions can be found in Platt [21], Smola and Schölkopf [22], Burges [7] and Keerthi and Shevade [23].

In order to solve the two Lagrange multipliers α_1, α_2, SMO first computes the constraints on these multipliers and then solves for the constrained minimum. For convenience, all quantities that refer to the first multiplier will have a subscript 1, while all quantities that refer to the second multiplier will have a subscript 2. The new values of these multipliers must lie on a line in (α_1, α_2) space, and in the box defined by $0 \leqslant \alpha_1, \alpha_2 \leqslant C$:

$$\alpha_1 y_1 + \alpha_2 y_2 = \alpha_1^{\text{old}} y_1 + \alpha_2^{\text{old}} y_2 = \text{constant}. \tag{20}$$

Without loss of generality, the algorithm first computes the second Lagrange multipliers α_2^{new} and successively uses it to obtain α_1^{new}. The box constraint $0 \leqslant \alpha_1, \alpha_2 \leqslant C$, together with the linear equality constraint $\Sigma \alpha_i y_i = 0$, provides a more restrictive constraint on the feasible values for α_2^{new}. The boundary of feasible region for α_2 can be applied as follows:

$$\text{If } y_1 \neq y_2;\ L = \max(0, \alpha_2^{\text{old}} - \alpha_1^{\text{old}}),\ H = \min(C, C + \alpha_2^{\text{old}} - \alpha_1^{\text{old}}),$$

$$\text{If } y_1 = y_2;\ L = \max(0, \alpha_1^{\text{old}} + \alpha_2^{\text{old}} - C),\ H = \min(C, C + \alpha_1^{\text{old}} + \alpha_2^{\text{old}}). \tag{21}$$

The second derivative of the objective function along the diagonal line can be expressed as

$$\eta = K(\mathbf{x}_1, \mathbf{x}_1) + K(\mathbf{x}_2, \mathbf{x}_2) - 2K(\mathbf{x}_1, \mathbf{x}_2). \tag{22}$$

Under normal circumstances, the objective function will be positive definite, there will be a minimum along the direction of the linear equality constraint, and η will be greater than zero. In this case, SMO computes the minimum along the direction of the constraint:

$$\alpha_2^{\text{new}} = \alpha_2^{\text{old}} + \frac{y_2(E_1^{\text{old}} - E_2^{\text{old}})}{\eta}, \tag{23}$$

where E_i is the prediction error on the ith training example. As a next step, the constrained minimum is found by clipping the unconstrained minimum to the ends of the line segment:

$$\alpha_2^{\text{new,clipped}} = \begin{cases} H & \text{if } \alpha_2^{\text{new}} \geqslant H, \\ \alpha_2^{\text{new}} & \text{if } L < \alpha_2^{\text{new}} < H, \\ L & \text{if } \alpha_2^{\text{new}} \leqslant L. \end{cases} \tag{24}$$

Now, let $s = y_1 y_2$. The value of α_1^{new} is computed from the new α_2^{new}:

$$\alpha_1^{\text{new}} = \alpha_1^{\text{old}} + s(\alpha_2^{\text{old}} - \alpha_2^{\text{new}}). \tag{25}$$

Solving Eq. (13) for the Lagrange multipliers does not determine the threshold b of the SVM, so b must be computed separately [21]. The following threshold b_1, b_2 are valid when the new α_1, α_2 are not at the each bounds, because it forces the output of the SVM to be y_1, y_2 when the input is $\mathbf{x}_1$, $\mathbf{x}_2$ respectively:

$$b_1 = E_1 + y_1(\alpha_1^{\text{new}} - \alpha_1^{\text{old}})K(\mathbf{x}_1, \mathbf{x}_1) + y_2(\alpha_2^{\text{new,clipped}} - \alpha_2^{\text{old}})K(\mathbf{x}_1, \mathbf{x}_2) + b^{\text{old}},$$

$$b_2 = E_2 + y_1(\alpha_1^{\text{new}} - \alpha_1^{\text{old}})K(\mathbf{x}_1, \mathbf{x}_2) + y_2(\alpha_2^{\text{new,clipped}} - \alpha_2^{\text{old}})K(\mathbf{x}_2, \mathbf{x}_2) + b^{\text{old}}.$$

When both b_1 and b_2 are valid, they are equal. When both new Lagrange multipliers are at bound and if L is not equal to H, then the interval between b_1 and b_2 are all thresholds that are consistent with the Karush–Kuhn–Tucker conditions which are necessary and sufficient conditions for an optimal point of a positive definite QP problem [7]. In this case, SMO chooses the threshold to be halfway between b_1 and b_2 [19].

6. Condition classification of butterfly valves

Fig. 8 shows the flow chart of the proposed procedure for detecting the cavitation of butterfly valve. Training of the SVM is carried out using the Platt's SMO algorithm. Both constant width RBF kernel, where the width σ is found by iteration, and the averaged width RBF kernel [9] is used as a kernel function. Types of SVM kernel function are used to compare the classification performance. The statistical features, namely, mean, standard deviation, shape factor and kurtosis are used to distinguish between normal (non-cavitating) and cavitating valve. Sequentially, it provides input vectors of SVM and SOFM [24,25] for training. We separate the total data set of 350 in a training data set of 100 (cavitation: 60, non-cavitation: 40) and a test data set of 250

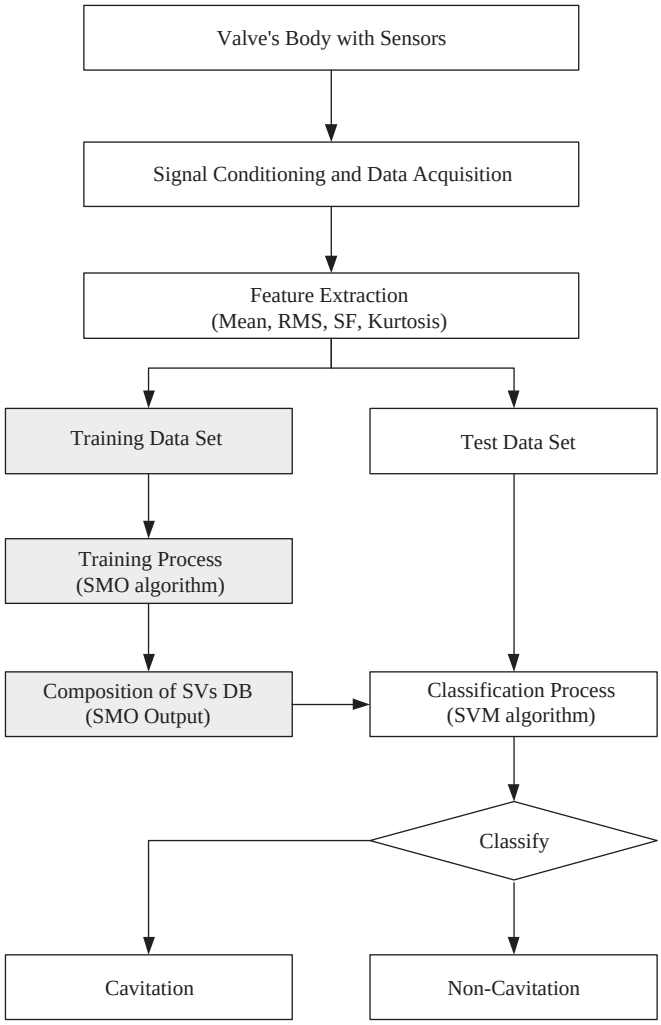

Fig. 8. Flow chart of classification system.

(cavitation: 150, non-cavitation: 100) as shown in Table 3. Thus, the size of training and test data is 100, 250 respectively.

The selection of RBF kernel width σ for good classification performance is one of the major issues in SVMs. The kernel width determines the radius of the hypersphere enclosing part of the data as a classifier boundary in a multidimensional feature spaces. If the value is too small, the enclosed feature space is also be very small leaving a significant part of the data that may lead to unsatisfactory classification. Also a very large value signifies a large enclosed feature space leading to an overlap between classes and possible misclassification [10]. An iterative process selecting an optimum value based on the full feature set determines the width of the Gaussian RBF kernel σ. Fig. 9 shows the influences of σ on the classification rate and the number of SVs. It can be seen that when σ takes a small value (e.g., $\sigma = 0.1$), testing error rate is large (horizontal 7.6%, vertical

Table 3
Composition for training and test data

	Condition	Valve opening (%)	No. of data
Training data	Cavitation	20	20
		30	20
		40	20
	Non-cavitation	50	20
		60	20
Test data	Cavitation	20	50
		30	50
		40	50
	Non-cavitation	50	50
		60	50

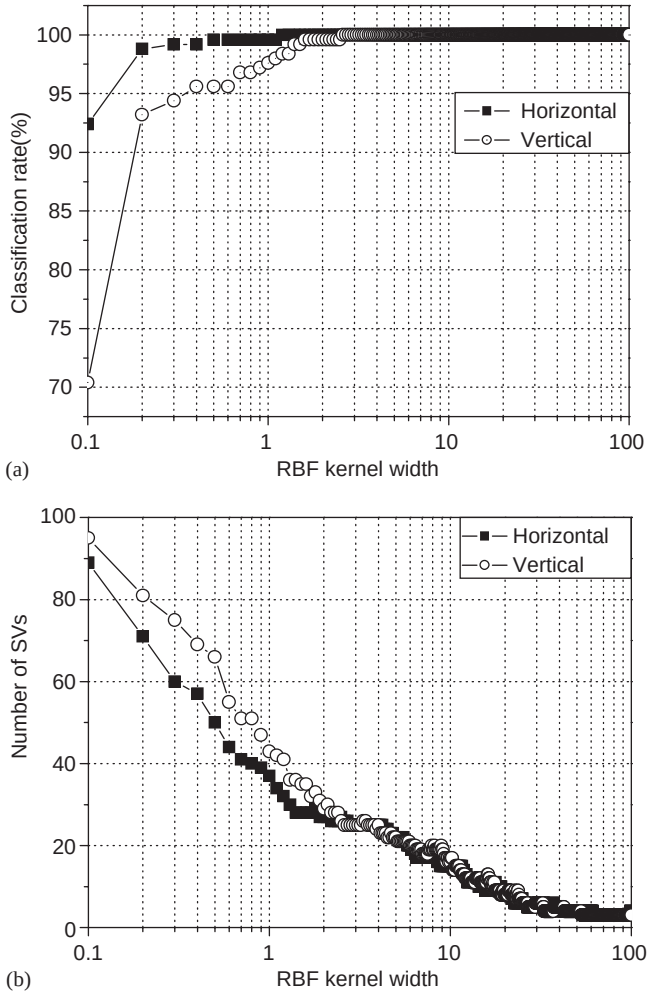

Fig. 9. Estimation of RBF parameter.

29.6%). Then, the testing error rate decreases with increment of σ. The number of SVs decreases with increment of σ and remains a minimum value of 3 during a certain range of σ (i.e., horizontal 61.6–97, vertical 52.9–95). Again, it tends to increase as σ continues with increment. The number of SVs can be considered as an indicator for the generalization ability of the SVM and also be used to give an upper bound of the generalization error of the hard margin SVM. The fewer the number of SVs the lower the upper bound of the generalization error [26,17]. From Fig. 9, an optimum value for best performance is $\sigma = 52.9$, 61.1 for horizontal and vertical directions respectively, while the number of SVs is all 3 for the two directions. Thus these values are selected for further experiments.

Another method for determining the width using the standard deviation was proposed by Jack and Nadi [9]. This method is to calculate the standard deviation of all the members of a given class for each feature in the input vector and determines an averaged value by taking the norm of vector formed from the individual standard deviations. Tables 4 and 5 illustrate the standard deviations of the input features and its averaged values for horizontal and vertical directions. As the distribution of each class is assumed Gaussian in nature, the averaged value for cavitation condition with non-Gaussian distribution become very large. This averaged value also used directly in Eq. (19).

The performance of SVMs and SOFM are shown in Table 6. In Table 6, we have listed the classification rate, RBF kernel width used and the number of SVs for training and test. The classification rate (%) is determined by the ratio of correct classifications on the whole training or test set, respectively. The averaged width SVM and constant width SVM with optimized value from Fig. 9 are the most classification rate, with a best training and test performance of all 100% for horizontal and vertical directions. But, the constant width SVM optimized has a smaller number of SVs than for averaged width SVM. The fewer the number of SVs the lower the upper

Table 4
Standard deviations of the input features for horizontal direction

Class	Standard deviation				Averaged value
	Mean	rms	Shape factor	Kurtosis	
Cavitation	0.002	0.176	4.867	0.116	1.290
Non-cavitation	0.016	0.044	5.248	54.185	14.873
Mean of standard deviation	0.009	0.110	5.056	27.150	8.082

Table 5
Standard deviations of the input features for vertical direction

Class	Standard deviation				Averaged value
	Mean	rms	Shape factor	Kurtosis	
Cavitation	0.001	0.320	10.939	0.076	2.834
Non-cavitation	0.013	0.084	4.692	38.918	10.927
Mean of standard deviation	0.007	0.202	7.815	19.497	6.880

Table 6
Performance comparisons of classifiers

Classifier	Direction	Classification rate (%)		RBF kernel width	Number of SVs
		Training	Test		
SVM (average width)	Horizontal	100	100	8.08	17
	Vertical	100	100	6.88	19
SVM (constant width)	Horizontal	100	100	52.9	3 (optimum)
	Vertical	100	100	61.1	3 (optimum)
	Horizontal	100	99.6	1.0	37
	Vertical	100	97.6	1.0	43
SOFM	Horizontal	100	87.2	—	—
	Vertical	100	84.8	—	—

Table 7
Classification rate and number of SVs according to different kernel functions

Kernel	Classification rate (%)				Number of SVs	
	Training		Test			
	Horizontal	Vertical	Horizontal	Vertical	Horizontal	Vertical
Linear	100	100	100	100	2	2
Polynomial ($d = 2$)	97	100	99.6	100	2	2
Gaussian RBF	100	100	100	100	3	2
χ^2	70	100	86	100	30	2

bound of the generalization error [17]. The number of SVs can be considered as an indicator for the generalization of the SVM. The SOFM achieves all 100% on training set and 87.2% and 84.8% on the test set for horizontal and vertical directions, respectively. These classification errors are due to the misclassification of cavitation data, i.e. some cavitation data for the valve opening of 20% were classified as those for no-condition condition.

Since it is not possible to determine a priori which kernel function works best for which data set, considerable time is spent on trying different kernel functions. However, it has to be chosen carefully since an inappropriate kernel can lead to poor performance. The kernels investigated were borrowed from the pattern recognition literature as shown in Table 2 [14,27]. The classification performance of SVMs using different kernel functions is summarized in Table 7. The order of polynomial kernel function is 2. The classification results show that the performance of linear and RBF kernel functions is the best among the four types of kernels.

7. Conclusions

This study presents a novel scheme for cavitation detection of butterfly valves used in large pumping stations based on two classifiers, SVM and SOFM. The stationary features of vibration

acceleration signals are extracted from statistical moments. The classifiers are trained, and then classify normal (non-cavitation) and cavitation conditions of valves. The test results proved that the trained classifier has the capability to detect the cavitation. The classification accuracy of SVMs was better than that of SOFM and offer 100% success rate both for training sets and test sets. This system can provide the potential use for real-time implementation leading to possible development of an automated valve condition monitoring and diagnostic system. Also, this is vital to a valve monitoring strategy since it can be used to detect problems and assist in the timely scheduling of repairs to the valve, before a severe failure occurs.

References

[1] H. Hassis, Noise caused by cavitating butterfly and monovar valves, *Journal of Sound and Vibration* 225 (3) (1999) 515–526.
[2] C.N. Anderson, Valves, in: R.L. Sanks (Ed.), *Pumping Station Design*, Butterworth-Heinemann, London, 1989 (Chapter 5).
[3] B. Carlson, Avoiding cavitation in control valves, *ASHRAE Journal* 43 (6) (2001) 58–63.
[4] V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, Berlin, 1995.
[5] V.N. Vapnik, On the uniform convergence of relative frequencies of events to their probabilities, *Soviet Mathematics: Doklady* 9 (1968) 915–918.
[6] V.N. Vapnik, An overview of statistical learning theory, *IEEE Transactions on Neural Networks* 10 (5) (1999) 988–999.
[7] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery* 2 (2) (1998) 955–974.
[8] S.R. Gunn, Support vector machines for classification and regression, Technical Report, University of Southampton, 1998.
[9] L.B. Jack, A.K. Nandi, Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms, *Mechanical Systems and Signal Processing* 16 (2–3) (2002) 373–390.
[10] B. Samanta, Gear fault detection using artificial neural networks and support vector machines with genetic algorithms, *Mechanical Systems and Signal Processing* 18 (3) (2004) 625–644.
[11] I. Guyon, N. Christianini, Survey of support vector machine applications, *Proceedings of NIPS'99 Special Workshop on Learning with Support Vectors*, 1999. http://www.clopinet.com/isabelle/Projects/SVM/applist.html.
[12] E.G. Milewski, *The Essentials of Statistics*, vol. 1, Research and Education Association, Piscataway, NJ, 1996.
[13] T. Kimura, T. Tanaka, K. Fujimoto, K. Ogawa, Hydrodynamic characteristics of a butterfly valve- prediction of pressure loss characteristics, *ISA Transactions* 34 (1995) 319–326.
[14] K.R. Müller, S. Mika, G. Rätsch, K. Tsuda, B. Schölkopf, An introduction to kernel-based learning algorithm, *IEEE Transactions on Neural Network* 12 (2) (2001) 181–201.
[15] B. Schölkopf, *Support Vector Learning*, Oldenbourg, Munich, 1997.
[16] B.S. Yang, D.S. Lim, J.L. An, Vibration diagnostic system of rotating machinery using artificial neural network and wavelet transform, in: *13th International Congress on COMADEM*, Houston, USA, 2000, pp. 923–932.
[17] D.J. Strauss, G. Steidel, Hybrid wavelet-support vector classification of waveforms, *Journal of Computational and Applied Mathematics* 148 (2002) 375–400.
[18] V.N. Vapnik, *Estimation of Dependences Based on Empirical Data*, Springer, Berlin, 1982.
[19] J. Platt, Fast training of support vector machines using sequential minimal optimization, in: B. Schölkopf, et al. (Eds.), *Advances in Kernel Methods—Support Vector Learning*, MIT Press, Cambridge, MA, 1999, pp. 185–208.
[20] E. Osuna, R. Freund, F. Girosi, Improved training algorithm for support vector machines, in: *Proceedings of IEEE Neural Networks for Signal Processing*, 1997, pp. 276–285.
[21] J. Platt, Sequential minimal optimization: a fast algorithm for training support vector machines, Technical Report 98-14, Microsoft Research, Redmond, Washington, 1998. http://www.research.microsoft.com/~jplatt/smo.ttml.

[22] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Technical Report NC2-TR-1998-030, 1998. http://www.neurocolt.com.
[23] S.S. Keerthi, S.K. Shevade, SOM algorithm for least squares SVM formulations, Control Division Technical Report CD-02-8, 2002. http://guppy.mpe.nus.edu.sg/~mpessk.
[24] T. Kohonen, Engineering application of the self-organizing feature map, *Proceedings of IEEE* 84 (10) (1996) 1358–1384.
[25] B.S. Yang, D.S. Lim, S.Y. Seo, M.H. Kim, Defect diagnostics of rotating machinery using SOFM and LVQ, in: *Proceedings of Seventh International Congress on Sound and Vibration*, Garmisch Partenkirchen, Germany, 2000, pp. 567–574.
[26] N. Cristianini, J. Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge, 2000.
[27] O. Chapelle, P. Haffner, V.N. Vapnik, Support vector machines for histogram-based image classification, *IEEE Transactions on Neural Networks* 10 (5) (1999) 1055–1064.