A Robust One-Class Support Vector Machine Using Gaussian-Based Penalty Factor and Its Application to Fault Detection

T. Prayoonpitak and S. Wongsu

Abstract—In recent years, one-class support vector machine (SVM) approaches have received particular attention in fault detection since only one class of the data is required for training. However, the training data can be corrupted with the outliers that influence classifier performance significantly. In this paper, a Gaussian-based penalisation has been proposed in the formation of a robust one-class SVM model which constructs the decision boundaries that are robust to the outliers without compromising the classification performance. The efficacy of the proposed method has been compared with the traditional one-class SVM and a previous robust one-class SVM method in the literature when applied in three datasets: the Iris's Fisher dataset, banana-shaped dataset and MFPT bearing fault dataset. It is shown that the proposed robust one-class SVM outperforms other methods.

Index Terms—Robust one-class SVM, penalty factor, fault detection, outliers.

I. INTRODUCTION

Novelty detection is the task of classifying test data that differ in some respect from the data that are available during training. This may be referred as “one-class classification”, in which a model is constructed to describe a dataset when only one class of dataset is available. The model is then used to detect whether a new sample resembles the training dataset according to the constructed description. Many methods have been proposed to solve one-class classification problems in recent years. Among all the method available for one-class classification problems, Support vector machine (SVM) is one of the most famous and popular techniques which has been applied for fault detection in industrial processes [1].

SVM is an advanced classification method from machine learning technique that has excellent generalisation ability compared with the traditional intelligent methods. Given two classes of data, the basic idea of SVM is to construct a hyperplane to separate two class datasets with a maximum margin [2]. In some cases, the datasets cannot be linearly separated. Thus, the kernel is used to project the datasets into a higher dimensional feature space and the datasets may be linearly classified more efficiently with SVM model. In order to apply SVM for fault detection, both the normal and the other faulty datasets should be collected for training SVM. However, the faulty datasets are difficult to be obtained

because it is expensive to make the faulty conditions and there are many and varied fault categories. Therefore, it is very common in practice that only the normal operation data set is available.

One-class SVM is a support vector method that needs only the normal operation dataset for training. One-class SVM determines the location of the decision boundary using only those data that lie closest to it. If there is a new data sample located within the decision boundary, it will be labelled as a normal data point. In contrast, it will be classified as an abnormality when it lies outside of the decision boundary.

In practical dataset, the training dataset can be corrupted by outliers or mislabelled data. It turns out that a one-class SVM is sensitive to such outliers in the training data. Thus, the new data may be incorrectly labelled by the decision boundary which constructed by the training data contaminated by outliers. To solve this problem, a robust one-class SVM is proposed in this paper. The important key concept of this method is the adaptive penalty factors which are designed to make outliers have less influences on the decision boundary of the one-class SVM [3]. Instead of computing the penalty factors using the 2-norm-based distance as proposed in [3], we propose to penalise the slack variables based on the weights that are a Gaussian function. We believe that this Gaussian-based penalty factor makes full use of the data distribution and provides more robust performance against the outliers.

The remainder of this paper is organised as follows. Section II briefly reviews the concept of one-class SVM and presents the detail of the proposed robust one-class SVM. Section III is devoted to present the simulation results. Section IV analyses the practical datasets and presents experimental results and finally Section V concludes the paper.

II. METHODS

A. One-Class SVM

The one-class SVM (OCSVM) proposed by Schölkopf *et al.* [4] is a support vector method that aims to solve the one-class classification problem. By treating the origin as a representative of outliers or faulty samples, the one-class classification problem is solved by formulating the decision boundary which has the maximum margin between the normal dataset and the origin [4]. Samples that fall outside this boundary are then classified as outliers or faulty samples.

Let $\mathbf{x}_i \in R^m, i = 1, \dots, N$ denote N samples of m -dimensional

Manuscript received January 30, 2016; revised June 8, 2016.

The authors are with the Department of Control Systems & Instrumentation Engineering, King Monkut's University of Technology Thonburi, Bangkok, Thailand (e-mail: teera.p@pttplc.com, sarawan.won@kmutt.ac.th).

normal data which are all associated to a class label $y_i = \{1\}$ and $\varphi(\mathbf{x}_i)$ is the mapping function that maps a data point from the input space to the feature space F . The one-class SVM optimisation problem can be formulated as follows:

$$\min_{\mathbf{w} \in F, \xi_i \in \mathbb{R}^N, \rho \in \mathbb{R}} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{N\nu} \sum_{i=1}^N \xi_i - \rho \quad (1)$$

subject to $\mathbf{w} \cdot \varphi(\mathbf{x}_i) \geq \rho - \xi_i, \quad \xi_i \geq 0$

where $\nu \in (0,1]$ corresponds to the upper bound of the fraction of the training data points that lie outside the decision boundary, ξ_i is the slack variable to relax the optimality constraints, while $\mathbf{w}$ and ρ are the vector and offset of the decision boundary in feature space, respectively.

Based on the Lagrangian and using the kernel trick, we can obtain the following dual problem of OCSVM

$$\min_{\alpha} \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (2)$$

subject to $0 \leq \alpha_i \leq \frac{1}{N\nu}, \quad \sum_{i=1}^N \alpha_i = 1$

where α_i represents the Lagrange's multiplier and all data points corresponding to nonzero α_i are called support vectors, and $K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$ is the kernel function.

One of the most common choices for the kernel function is the Gaussian kernel function because there exists only one parameter to be tuned [5]. The Gaussian kernel function is in the following form:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2}\right) \quad (3)$$

where σ is the width parameter to be selected.

By solving the optimisation problem in (2), we can obtain the decision function $f(\mathbf{x})$

$$f(\mathbf{x}) = \mathbf{w} \cdot \varphi(\mathbf{x}) - \rho = \sum_{i=1}^N \alpha_i K(\mathbf{x}_i, \mathbf{x}) - \rho \quad (4)$$

The data point will be labelled as normal if $f(\mathbf{x}) \geq 0$; otherwise, it is labelled as abnormal.

B. Robust One-Class SVM

In OCSVM problem, usually it is assumed that the training samples contain only positive or normal samples. However, in practice it is hard to satisfy this assumption. Not only because clean data are not always easy to be obtained as training samples, but also some outliers could be included in the training set due to mislabelling. To deal with these circumstances, S. Yin *et al.* [3] have proposed a robust one-class SVM technique which modifies the penalty factor $1/N\nu$ of the traditional OCSVM and is formulated as

$$\min_{\mathbf{w} \in F, \xi_i \in \mathbb{R}^N, \rho \in \mathbb{R}} \frac{1}{2} \|\mathbf{w}\|^2 + \theta \sum_{i=1}^N \hat{d}_i \xi_i - \rho \quad (5)$$

subject to $\mathbf{w} \cdot \varphi(\mathbf{x}_i) \geq \rho - \xi_i, \quad \xi_i \geq 0$

where θ is a constant and $\hat{d}_i$ ($i = 1, \dots, N$) are the adaptive

penalty factors related to the distance between $\mathbf{x}_i$ and the centre of the dataset.

The penalty factor $\hat{d}_i$ is calculated as

$$\hat{d}_i = \frac{d_{i,\max}}{d_i} \quad (6)$$

$$d_i = \|\mathbf{x}_i - \mathbf{x}_c\|^2 \quad (7)$$

where $d_{i,\max}$ is the maximum distance between the point $\mathbf{x}_i$ and the centre of the dataset $\mathbf{x}_c$.

The centre $\mathbf{x}_c$ is calculated by the tSL-centre technique [6], [7] which is robust to noise and outliers and it is expressed as

$$\mathbf{x}_c = \sum_{i=1}^N k_i \mathbf{x}_i \quad (8)$$

In which

$$k_i = \frac{1}{\sqrt{1 + 4\|\mathbf{x}_i\|^2}} \bigg/ \sum_{j=1}^N \frac{1}{\sqrt{1 + 4\|\mathbf{x}_j\|^2}} \quad (9)$$

According to (6), the influence of outliers in the formation of the decision boundary is reduced by applying a small penalty factor to samples which are far away from the centre so that the outliers will be forced to fall outside the boundary.

Because of its minor modification to the traditional OCSVM formulation, this robust one-class SVM can be easily incorporated in the original solver.

Even though the simulation results in [3] have provided positive evidence of this technique in dealing with outliers, its lack of control in the shape or the rate of decay of the penalty function may cause problems when dealing with dataset with different distribution and severity of outlier contamination.

C. Proposed Robust One-Class SVM

To provide a control of the rate of decay of penalisation, we propose to use the weighted kernel function to calculate the penalty factors.

The weighted Gaussian function is calculated as

$$\hat{d}_i^k = d_i^{\text{weight}} \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_c\|^2}{\sigma_d^2}\right) \quad (10)$$

$$d_i^{\text{weight}} = \frac{d_{i,\max} - d_i}{d_{i,\max} - d_{i,\min}} \quad (11)$$

where σ_d is the parameter to control the width of the kernel, $d_i^{\text{weight}} \in [0,1]$ are the weights to scale the resulting penalty factors to range between 0 and 1 and force zero weight to the farthest sample, d_i is the distance defined in (7) and $d_{i,\min}$ is the minimal distance.

The concept of using the Gaussian to calculate the penalty factors is the same as the distance-based penalty function in (6), i.e. the outliers which are far away from the centre are penalised with small weights and are likely to be outside of the boundary. However, unlike the distance-based penalty function, the kernel width σ_d allows us to control the rate of penalisation or the decaying rate of the penalty factors as a

function of the distance, as Fig. 1 illustrates. Note that the distance-based penalty factors plotted in Fig. 1 are scaled to range from 0 to 1 so that they can be superimposed with the Gaussian penalty factors. Fig. 2 displays the surface plots of the Gaussian penalty factor with different Gaussian widths, overlaying on a scatter plot of samples drawn from a two-dimensional Gaussian with zero mean and covariance matrix $\Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 5 \end{bmatrix}$.

Fig. 1. Distanced-based and Gaussian-based penalty factors as a function of 2-norm distance between the samples and the centre of the data.

Fig. 2. Gaussian-based penalty factor surface for normal distributed dataset (a) width = 1 (b) width = 0.38.

It can be seen that the value of Gaussian width has a great influence on the distribution of the penalty factors and, as a result, affects the final outcome of this robust one-class SVM. In this work, we propose to estimate the Gaussian width of penalty factors as a function the eigenvalues of the PCA-projected data. The steps of proposed Gaussian-based robust one-class SVM can be summarised as follows:

- 1) The original training samples are scaled to zero mean and unitary standard deviation. The scaled data are preprocessed by PCA (principal component analysis) which decomposes the scaled training matrix $\mathbf{X} \in R^{N \times m}$ as

$$\mathbf{X} = \mathbf{t}_1 \mathbf{p}_1^T + \mathbf{t}_2 \mathbf{p}_2^T + \dots + \mathbf{t}_l \mathbf{p}_l^T + \mathbf{E} \quad (12)$$

where $\mathbf{t}_i$ is the scores vector, $\mathbf{p}_i$ is the loading vector, l is the number of principal components (PCs) retained in PC model, and $\mathbf{E}$ is the residual matrix. The loading vector $\mathbf{p}_i$ can be computed as the i th eigenvector of the covariance matrix $\text{cov}(\mathbf{X})$ and λ_i is the eigenvalue corresponding to the eigenvector $\mathbf{p}_i$. The scores vector $\mathbf{t}_i \in R^l$ represents the projections of the scaled data sample $\mathbf{x}$ along the PCs. The score matrix $\mathbf{T}$ corresponding to the projections of the scaled training matrix $\mathbf{X}$ can be calculated as:

$$\mathbf{T} = \mathbf{X}[\mathbf{p}_1 \quad \mathbf{p}_2 \quad \dots \quad \mathbf{p}_l], \mathbf{T} \in R^{N \times l} \quad (13)$$

- 2) The robust one-class SVM model is developed by

solving the optimisation problem (5) with $\hat{d}_i$ being replaced by $\hat{d}_i^k$ obtained from (10). The Gaussian width σ_d should be related to the spread of the samples in the training set. It is estimated as the average of the square roots of the eigenvalues of the retained PCs, i.e.

$$\sigma_d = \frac{1}{l} \sum_{i=1}^l \sqrt{\lambda_i} \quad (14)$$

- 3) In the testing process, the test sample is first normalised by the mean and standard deviation of training samples, and then evaluated by the built model to obtain its predicted label.

III. SIMULATION RESULTS

In this section, we conduct two groups of experiments to verify the performance of the one-class SVM and the robust one-class SVMs. The Iris' Fisher dataset and the banana data set are selected to perform the experiments and compare the accuracies of the traditional one-class SVM, the robust one-class SVM proposed by [3] (ROCSVM), and our Gaussian-based robust one-class SVM model.

A. Case I: Iris' Fisher Dataset

The data that we used in this experiment are obtained from Iris' Fisher dataset which contains 50 samples of 3 types of Iris flower. Two species, the Iris versicolor and Iris virginica with petal length and petal width feature, are selected to perform the experiment.

Here we select the Iris versicolor as the positive class (normal class), and the virginica species as the negative class. From the scatter plot in Fig. 3, we can clearly see that the two classes are quite poorly-separated.

We execute 50 runs of experiments. For each run, thirty samples of positive class are included in the training set and the remaining twenty samples are for testing. To evaluate the robustness of the considered techniques against outliers, 10% (3 samples) and 20% (6 samples) of negative samples are randomly picked and contaminated in the training set.

Fig. 3. Scatter plot of the features selected Iris' Fisher dataset before the PCA preprocessing.

Twenty samples of the remaining negative data points are included in the test set. Therefore, in the test set there are 20 samples from the positive class and 20 samples from the negative class. All data are normalised before training and testing as described in the method section.

For one-class SVM, the regularisation parameter $\nu = 0.05$. The constant θ of all robust one-class SVM techniques is set

to $1/N\nu$ to make an equivalent comparison. For each run, the width parameter σ for Gaussian kernel is varied from 0.5 to 8.0 with step size of 0.5 and the corresponding classification accuracy when tested on the test set is computed. After 50 executions, we select the optimal kernel parameter for each SVM technique from the one that gives the best performance, i.e. the highest average accuracy. The best average accuracies obtained from all SVM models are listed in Tables I and II, where standard deviations are given in brackets. We also analyse the difference between the proposed method and the others by the Wilcoxon signed-ranks test [8]. The last two rows in the tables list the test results and their p -values, where $h = 1$ indicates that there is significant difference between the proposed method and the other, while $h = 0$ indicates when there is not.

TABLE I: PERCENT ACCURACIES FOR IRIS'S FISHER DATASET WITH $\nu = 0.05$, OUTLIERS = 10%

Dataset	OCSVM	ROCSVM	Proposed
Iris' Fisher	79.05(7.16)	79.05(7.16)	86.40(5.28)
p -Value	1.20×10^{-7}	1.20×10^{-7}	-
h	1	1	-

TABLE II: PERCENT ACCURACIES FOR IRIS'S FISHER DATASET WITH $\nu = 0.05$, OUTLIERS = 20%

Dataset	OCSVM	ROCSVM	Proposed
Iris' Fisher	73.25(7.46)	73.25(7.46)	87.60(6.10)
p -Value	1.89×10^{-9}	1.89×10^{-9}	-
h	1	1	-

Fig. 4. Scatter plots of contaminated training set of the Iris data and the resulting decision boundaries of (a) OCSVM, (b) ROCSVM, and (c) the proposed method using the optimal kernel widths.

To compare the performances of all considered SVM methods when the training data are contaminated by outliers, their decision boundaries for 10% contamination case are shown in Fig. 4. The proposed robust one-class SVM results are in good performances on this dataset. It apparently shows that the outliers fall outside of the decision boundary. In contrary, the decision boundaries of one-class SVM and ROCSVM are loose and shifted towards the outliers.

From the results in Tables I and II, it is apparent that the proposed method is superior to the other methods on all cases of Iris' Fisher dataset. Specifically, among the three methods, the proposed method obtains the highest accuracies.

B. Case II: Banana Dataset

In this experiment, we perform experiment on the dataset which is distributed in banana shape. The banana-shaped data are selected to investigate the performance of the considered SVM techniques in non-Gaussian distributed data.

Fig. 5. Scatter plot of the banana dataset and outliers before the PCA preprocessing.

The banana-shaped dataset contains 150 samples of positive samples and 100 samples of Gaussian-distributed negative class. In Fig. 5, the scatter plot of positive and negative samples for this dataset shows that the two classes are clearly separable.

The training data are composed of 100 randomly selected positive samples contaminated with 10% (10 samples) and 20% (20 samples) from the negative class. To construct test data, 50 samples from the banana dataset and 50 samples from the Gaussian ones which are not already included in the training set are randomly selected.

For one-class SVM, the regularisation parameter $\nu = 0.05$ and a constant $\theta = 1/N\nu$ is used for the two robust SVM methods. The width parameter for Gaussian kernel is varied from 0.5 to 8.0, with step size of 0.5 and the corresponding classification accuracy for the testing samples is computed. Tables III and IV show the best accuracies averaging from 50 runs with standard deviations and the results of the Wilcoxon signed-ranks test also being displayed.

TABLE III: PERCENT ACCURACIES FOR BANANA DATASET WITH $\nu = 0.05$, Outliers = 10%

Dataset	OCSVM	ROCSVM	Proposed
Banana	62.98(7.80)	61.78(6.87)	89.18(3.41)
p -Value	7.46×10^{-10}	7.47×10^{-10}	-
h	1	1	-

To illustrate the performance of the three methods on the banana-shaped dataset, we plot their decision boundaries in Fig. 6. The decision boundaries of one-class SVM method and ROCSVM are apparently loose; the decision boundaries

in Fig. 6(a) and 6(b) are likely to shift towards the outliers. It is clearly shown in Fig. 6(c) that the method we proposed is capable of describing the banana-shaped dataset contaminated with outliers accurately

TABLE IV: PERCENT ACCURACIES FOR BANANA DATASET WITH $\nu = 0.05$, OUTLIERS = 20%

Dataset	OCSVM	ROCSVM1	Proposed
Banana	57.28(5.06)	55.80(4.73)	90.80(3.02)
p -Value	7.36×10^{-10}	7.36×10^{-10}	-
h	1	1	-

Fig. 6. Scatter plots of contaminated training set of the banana data and the resulting decision boundaries of (a) OCSVM, (b) ROCSVM, and (c) the proposed method using the optimal kernel widths.

From the results in Tables III and IV, it is apparent that the proposed method is superior to the other methods for all levels of outliers. Specifically, it can be observed that the Gaussian-based robust one-class SVM performs equally well for 10% and 20% outlier contamination cases with accuracies approximately 90%. On the contrary, the OCSVM and ROCSVM greatly suffer from the increasing level of outlier contamination.

IV. APPLICATION TO MFPT BEARING FAULT DATASET

In this section, the proposed technique is tested on the

bearing fault dataset to illustrate its applicability in practice. The comparison with other two methods, i.e. the OCSVM and ROCSVM, is also made

A. Experiment Dataset

The bearing data were obtained from the Society for Machinery Failure Prevention Technology (MFPT) [9]. The time-domain vibration signals of bearing were collected from the normal case, outer race fault case, and the case of inner race fault with input shaft rate of 25Hz [10]. The vibration signal collected from different fault conditions and different sampling rate were divided into non-overlapping 0.15-second width windows. The details of this dataset are listed in Table V.

TABLE V: DETAILS OF MFPT BEARING FAULT DATASET

Condition class	Number of samples	Sampling frequency (Hz)	Data length (samples)
Normal condition	120	97.6	14,648
Outer race fault	120	97.6	14,648
Inner race fault	140	48.8	7,324
Other Outer race fault	140	48.8	7,324

Bearing enveloped analysis which is a well-known signal processing technique for bearing fault detection was applied to extract the fault features. We used the envelope kurtosis with frequency bands 1 kHz to 4.05 kHz and 1 kHz to 9.1 kHz to construct a set of two-dimensional features representing the three fault conditions.

Fig. 7. Scatter plot of features of MFPT dataset before the PCA preprocessing.

Scatter plot of features in Fig. 7 shows that the normal dataset is clearly separable from the other two faulty datasets, while the outer race and inner race fault data are poorly separated.

In this experiment, we consider three cases as follows:

Case I: Normal as positive class and inner race faulty as negative class

60 samples of normal condition of bearing are randomly selected in the training dataset which is contaminated with 10% (6 samples) and 20% (12 samples) of inner race fault data. The testing data is constructed by randomly selected 60 data points from normal condition of bearing and 60 data points from inner race fault bearing.

Case II: Inner race fault as positive class and outer race faulty as negative class

60 samples of inner race fault data are randomly selected in the training dataset which is contaminated with 10% (6 samples) and 20% (12 samples) of outer race fault data. The testing data is constructed by randomly selected 60 data points from inner race fault condition and 60 data points from outer race fault bearing.

Case III: Outer race fault as positive class and inner race fault as negative class:

60 samples of outer race fault data are randomly selected in the training dataset which is contaminated with 10% (6 samples) and 20% (12 samples) of inner race fault data. The testing data is constructed by randomly selected 60 data points from outer race fault condition and 60 data points from inner race fault bearing.

In each case, the regularisation parameter ν is set at 0.05 for one-class SVM and $\theta = 1/N\nu$ for all robust one-class SVM. Again, the optimal kernel widths are obtained by performing the grid search over the range of 0.5 – 8.0 and considering the average accuracy over 50 executions as the performance index of the search. Case I represents the case where the contaminated outliers are far away from the positive case.

B. Result

The results of Case I are listed in Tables VI and VII. Overall, the average accuracy of the proposed method is higher than accuracies of the other two methods for both 10% and 20% contamination levels.

TABLE VI: PERCENT ACCURACIES FOR CASE I WITH $\nu = 0.05$, OUTLIERS = 10%

Dataset	OCSVM	ROCSVM	Proposed
Bearing	89.78(4.23)	89.78(4.23)	96.32(1.99)
p -Value	5.14×10^{-9}	5.14×10^{-9}	-
h	1	1	-

TABLE VII: PERCENT ACCURACIES FOR CASE I WITH $\nu = 0.05$, OUTLIERS = 20%

Dataset	OCSVM	ROCSVM	Proposed
Bearing	80.50(6.28)	80.50(6.28)	96.48(2.26)
p -Value	7.49×10^{-10}	7.49×10^{-10}	-
h	1	1	-

Fig. 8. Scatter plots of contaminated training set of Case I bearing dataset and the resulting decision boundaries of (a) ROCSVM, and (b) the proposed method using the optimal kernel widths.

Fig. 8(a) shows that the decision boundary of ROCSVM is tight and the ROCSVM mistakes the outliers as positive classes. The increased number of falsely identified positive samples deteriorates the overall accuracy as a result.

For Cases II and III, we explore the situations where the

outliers and the positive samples are close which make it difficult to define an appropriate decision boundary. The results of both cases are shown in Tables VIII – XI with corresponding boundaries plotted in Figs. 9 and 10. It can be clearly observed that even in such difficult cases where the contaminated outliers are poorly-separated from the actual positive samples; the proposed method can still provide satisfying results.

Fig. 9. Scatter plots of contaminated training set of Case II bearing dataset and the resulting decision boundaries of (a) ROCSVM, and (b) the proposed method using the optimal kernel widths.

TABLE VIII: PERCENT ACCURACIES FOR CASE II WITH $\nu = 0.05$, OUTLIERS = 10%

Dataset	OCSVM	ROCSVM1	Proposed
Bearing	70.20(5.10)	70.20(5.10)	86.10(3.41)
p -Value	7.47×10^{-10}	7.47×10^{-10}	-
H	1	1	-

TABLE IX: PERCENT ACCURACIES FOR CASE II WITH $\nu = 0.05$, OUTLIERS = 20%

Dataset	OCSVM	ROCSVM1	Proposed
Bearing	63.95(5.68)	63.95(5.68)	81.57(5.02)
p -Value	7.98×10^{-10}	7.98×10^{-10}	-
h	1	1	-

TABLE X: PERCENT ACCURACIES FOR CASE III WITH $\nu = 0.05$, OUTLIERS = 10%

Dataset	OCSVM	ROCSVM1	Proposed
Bearing	69.48(4.74)	69.48(4.74)	83.78(3.47)
p -Value	7.49×10^{-10}	7.49×10^{-10}	-
h	1	1	-

TABLE XI: PERCENT ACCURACIES FOR CASE III WITH $\nu = 0.05$, OUTLIERS = 20%

Dataset	OCSVM	ROCSVM1	Proposed
Bearing	63.65(5.81)	63.65(5.81)	77.97(5.17)
p -Value	1.29×10^{-9}	1.29×10^{-9}	-
h	1	1	-

In general, the Gaussian-based penalty factors are more robust to the samples that are far from the centre of the data points and try to force them to be outside the decision boundary. However, it can also misidentify the actual positive samples which happen to be located further from the centre.

Having said that, from all of the investigated cases which represent both close and far outliers we may conclude that the proposed penalty function has good potential to be used in practice.

Fig. 10. Scatter plots of contaminated training set of Case III bearing dataset and the resulting decision boundaries of (a) ROCSVM, and (b) the proposed method using the optimal kernel widths.

V. CONCLUSIONS

In this paper, we study the robustness of one-class SVM and its application to fault detection. We propose a novel method of penalisation that can decrease the influences of outliers by using adaptive penalty factor. The robust one-class SVM presented in literature (ROCSVM) solves the problem by adjusting the penalty factor based on the distance between the data point and the centre coordinate of the data set. However, this method is unable to make full use of the data distribution information. Therefore, the Gaussian function which makes use of the distribution of the data is proposed to adjust the penalty factor.

Then we test the performances of methods using three groups of experiments: Iris' Fisher dataset, banana-shaped dataset, and MFPT bearing fault dataset which is a practical industry dataset. Outliers are apparently contaminated in the training data to evaluate the robustness of each method. As shown by the experimental results, the proposed robust one-class SVM can generate the suitable decision boundaries and perform much better than other methods.

However, currently the width parameter for Gaussian kernel is exhaustively searched by using the grid search for the sake of fair comparison which is clearly not possible in practice. Future research will be focused on the study of the

parameter selection for Gaussian kernel when being incorporated with the proposed robust one-class SVM.

REFERENCES

- [1] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Processing*, vol. 99, pp. 215-249, June 2014.
- [2] B. Scholkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 3, no. 7, pp. 1443-1471, July 2001.
- [3] S. Yin, X. Zhu, and C. Jing, "Fault detection based on a robust one class support vector machine," *Neurocomputing*, vol. 145, pp. 263-268, December 2014.
- [4] B. Scholkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt, "Support vector method for novelty detection," *Advances in Neural Information Processing Systems*, vol. 13, pp. 582-588, 2000.
- [5] S. S. Keerthi and C. J. Lin, "Asymptotic behaviors of support vector machines with gaussian kernel," *Neural Computation*, vol. 15, no. 7, pp. 1667-1689, July 2003.
- [6] M. Liu, B. C. Vemuri, S.-I. Amari, and F. Nielsen, "Total bregman divergence and its applications to shape retrieval," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 3463-3468.
- [7] S. Yin and G. Wang, "A modified partial robust M-regression to improve prediction performance for data with outliers," in *Proc. IEEE International Symposium on Industrial Electronics*, 2013, pp. 1-6.
- [8] J. Demsar, "Statistical comparisons of classifiers over multiple data sets," *The Journal of Machine Learning Research*, vol. 7, pp. 1-30, 2006.
- [9] E. Bechhoefer. Condition based maintenance fault database for testing of diagnostic and prognostics algorithms. [Online]. Available: <http://www.mfpt.org/FaultData/FaultData.htm>
- [10] E. Bechhoefer, P. Menon, and M. Kingsley, "Bearing envelope analysis window selection using spectral kurtosis techniques," in *Proc. IEEE Conference on Prognostics and Health Management*, Montreal, 2011, pp. 1-6.

T. Prayoonpitak received the BEng degree in electrical engineering from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, 2007. He is currently an MS degree student in King Mongkut's University of Technology Thonburi, Bangkok, Thailand. His research interests include fault classification and machine learning.

S. Wongsas received the BEng degree in control systems and instrumentation from King Mongkut's University of Technology Thonburi, Bangkok, Thailand in 1998, and the MSc and PhD degrees in automatic control and systems engineering from Sheffield University, United Kingdom, in 2002 and 2007, respectively. She is currently an assistant professor in the Department of Control Systems and Instrumentation Engineering at King Mongkut's University of Technology Thonburi. Her research interests include intelligent fault detection and diagnosis.