Research Journal of Applied Sciences, Engineering and Technology 5(16): 4066-4070, 2013
DOI:10.19026/rjaset.5.4626
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: March, 16, 2012 Accepted: January 11, 2013 Published: April 30, 2013

**Research Article**

# A Fault Detection Model of Marine Refrigerated Containers

Jun Ji and Houde Han
Merchant Marine College, Shanghai Maritime University, Shanghai 201306, China

**Abstract:** A fault detection model based on One-Class Support Vector Machine was established to solve the large difference in sample size between the normal data and fault data of refrigerated containers. During the model training process, only the normal samples were needed to be learned, and an accurate identification of abnormal was achieved, which may solve the problem of lack of fault samples in practice. By comparison experiments between different kernel functions and kernel parameter optimization, a fault detection model of refrigerated containers based on One-Class Support Vector Machine was established, and the test results show that the model has a high recognition rate against abnormal of 97.4% and zero false alarm rate.

**Keywords:** Fault detection, one-class svm, parameter optimization, refrigerated container

## INTRODUCTION

Typically, equipment fault detection is dealt with as a pattern recognition issue of two type classification, however, there is little or no abnormal samples but a lot of normal samples in practice. Due to the lack of abnormal training samples, commonly used classification methods can not achieve good results, so setting alarm thresholds of a certain parameter that can reflect the equipment operation status is used in fault detection. In fact, only use one parameter is not accurate, because it is likely to result in miss alarm and false alarm. For example, though supply air temperature of refrigerated containers may reflect the operation conditions of the refrigeration unit most of time, when some faults happen, the supply air temperature is still within the normal range, while other parameters such as the exhaust air temperature of compressor has changed a lot, if threshold alarm method is yet used, miss alarm will take place.

Support Vector Machine (SVM) is a machine learning method with good generalization performance and outstanding ability of dealing with small samples, which was put forward on the basis of statistical learning by Vapnik (2000) in the 1990s. As a pattern classification method, SVM has aroused extensive attention recently. Based on structural risk minimization principle instead of the traditional empirical risk minimization principle, SVM introduced the kernel function method and then the classification problem come down to a quadratic programming problem, which has not only effectively overcome the high dimension and local minimization problems but also solved the nonlinear classification problem (Burges, 1998; Schwenker, 2000). Therefore, SVM has incomparable advantages in pattern classification. So far, SVM has been put into use in regression analysis, function estimates, isolated handwritten character recognition, web pages or automatic text categorization, face detection, computer intrusion detection, gene classification (Clarke *et al*., 2005; Liu *et al*., 2003; Dong *et al*., 2005; Sun *et al*., 2002; Osuna and Freund, 1997; Fugate and Gattikar, 2003; Guyon *et al*., 2002), etc.

The lack of abnormal samples of refrigerated containers determines its fault detection can not follow the conventional way and new ways and methods are needed. New methods can only use a large number of existing normal samples to achieve the early identification of abnormal conditions and failures. One-Class SVM is a new kind of SVM, which only need one class of samples as training samples, through adaptive learning of their distribution, effective recognition of different modes and states may be realized (Zhong and Cai, 2006). One-Class SVM was introduced to condition assessment of refrigerated containers, and the fault detection model of marine refrigerated containers based on One-Class SVM was established.

In this study, a fault detection model based on One-Class Support Vector Machine was established to solve the large difference in sample size between the normal data and fault data of refrigerated containers. During the model training process, only the normal samples were needed to be learned, and an accurate identification of abnormal was achieved, which may solve the problem of lack of fault samples in practice. By comparison experiments between different kernel functions and kernel parameter optimization, a fault detection model of refrigerated containers based on

**Corresponding Author:** Jun Ji, Merchant Marine College, Shanghai Maritime University, Shanghai, China

One-Class Support Vector Machine was established, and the test results show that the model has a high recognition rate against abnormal of 97.4% and zero false alarm rate.

## ONE-CLASS SVM CLASSIFICATION ALGORITHM

**One-class SVM classification principle:** One-Class SVM is originally used as high-dimensional distribution estimation (Scholkopf *et al.*, 2001) and one class data classification problem is proposed on the basis by Tax and others. First, the original samples were projected to high dimensional feature space through kernel mapping. Then, the distribution range model of study samples was established in the feature space, and the distribution area was asked to cover the training sample as compact as possible to construct the classification decision function. When unknown samples fall into the decision area, they are judged as normal. Otherwise they are judged as abnormal (Tax and Duin, 1999).

Suppose $x_i \in R$ is the training vector without any classification information and $\Phi$ is the corresponding mapping of the kernel function. Map x to a high dimensional feature space H. The following quadratic programming problem is needed to be solved in order to separate most of the samples with another type of samples (original point) with the largest interval:

$$\begin{cases} \min\limits_{w,b,\xi,\rho} \quad \frac{1}{2}\|w\|^2 + \frac{1}{vl}\sum_{i=1}^{l}\xi_i - \rho \\ s.t. \quad (w\cdot\varphi(x_i)) \ge \rho - \xi_i \\ \qquad \xi_i \ge 0, \quad i = 1,\ldots\ldots l \end{cases} \tag{1}$$

$v \in (0,1]$, its function is similar to the corresponding parameter in v-SVM. $\xi_i$ is a nonzero relaxation variable, which is used to punish the samples that cannot be completely separated. The final decision function has the following form:

$$f(x) = sign((w\cdot\varphi(x_i)) - \rho) \tag{2}$$

For most of the samples in the training set, they should satisfy $f(x_i) > 0$ if they are placed on the correct side of decision surface. However, this will make $\|w\|$ too big, so that the biggest classification interval $\rho/\|w\|$ will become smaller. In order to comprise between the two, parameter v is used to adjust.

To solve the above optimization function, the Lagrange coefficient $\alpha_i \geq \beta_i \geq 0$ are introduced to construct the Lagrange function:

$$L(w,\xi,\rho,\alpha,\beta) = \frac{1}{2}\|w\|^2 + \frac{1}{vl}\sum_{i=1}^{l}\xi_i - \rho - \sum_{i=1}^{l}\alpha_i((w\cdot\varphi(x_i)) - \rho + \xi_i) - \sum_{i=1}^{l}\beta_i\xi_i \tag{3}$$

Solve the partial differential equations for w $\xi$, $\rho$ respectively, and make the value of the equations be zero, then we have:

$$\begin{cases} w = \sum_{i=1}^{l}\alpha_i\varphi(x_i) \\ \alpha_i = \frac{1}{vl} - \beta \\ \sum_{i=1}^{l}\alpha_i = 1 \end{cases} \tag{4}$$

Substitute Eq. 4 into Eq. 2, the decision function is as follows:

$$f(x) = sign(\sum_{i=1}^{l}\alpha_i k(x,x_i) - \rho) \tag{5}$$

The dual form of the above optimization problem can be written as:

$$\begin{cases} \min \quad \frac{1}{2}\sum_{i=1}^{l}\alpha_i\alpha_j k(x_i,x_j) \\ s.t. \quad 0 \le \alpha_i \le \frac{1}{vl} \\ \qquad \sum_{i=1}^{l}\alpha_i = 1 \end{cases} \tag{6}$$

**One-class SVM anomaly detection algorithm:** The idea that using hyper spheres instead of hyper planes to divide the data, which was proposed by Tax has changed the data set description. Suppose the sample set X ($\boldsymbol{x}_i \in R$, $\boldsymbol{i} = 1, \ldots, l$) corresponds to the normal state, the so-called anomaly detection is to find a data set covering X and construct a decision function. For each sample, there is the relationship:

$$\begin{cases} f(x) \ge 0, \quad x \in C(x) \\ f(x) < 0, \quad x \notin C(x) \end{cases} \tag{7}$$

Then, the samples falling within C(x) will be determined as normal, otherwise determined as abnormal.

First, project the sample set into high dimensional feature space through kernel function. In order to reduce false rate, we should find a compressed sphere in the feature space that contains as much training samples as possible, which is called hyper sphere. Then introduce the relaxation variable $\xi_i$ to make the training samples included as much as possible in the hyper sphere during the guarantee that the hyper sphere is most compressed. This problem can be expressed as an optimization problem:

$$\begin{cases} \min\limits_{R,\xi_i,c} \quad R^2 + \frac{1}{vl}\sum_{i=1}^{l}\xi_i \\ s.t. \quad \|\varphi(x_i) - c\|^2 \le R^2 + \xi_i \\ \qquad \xi_i \ge 0, \quad i = 1,\dots\dots l \end{cases} \tag{8}$$

R for hyper sphere radius, c for circle center, v ∈(0,1]

In order to compromise between the hyper sphere radius and the number of the training samples it contains, when v is small, restrict the samples in the sphere to the greatest extent and when v is large, compress the size of the sphere as much as possible. Lagrange function is utilized to solve this optimization problem.

$$L(R, c, \xi_i, \alpha_i, \beta_i) = R^2 + \frac{1}{vl}\sum_{i=1}^{l}\xi_i - \sum_{i=1}^{l}\alpha_i(R^2 + \xi_i - \|\varphi(x_i) - c\|^2) - \sum_{i=1}^{l}\beta_i\xi \tag{9}$$

Seek the partial differential equations for R, c and ξ respectively, and make them equal to zero.

Dual form of this optimization problem is as follows:

$$\begin{cases} \sum_{i=1}^{l}\alpha_i = 1 \\ c = \sum_{i=1}^{l}\alpha_i x_i \\ 0 \le \alpha_i \le \frac{1}{vl} - \beta_i \le \frac{1}{vl} \end{cases} \tag{10}$$

$$\begin{cases} \min \quad \sum_{i=1}^{l}\alpha_i\alpha_j k(x_i, x_j) - \sum_{i=1}^{l}\alpha_i k(x_i, x_i) \\ s.t. \quad 0 \le \alpha_i \le \frac{1}{vl}, \quad \sum_{i=1}^{l}\alpha_i = 1 \end{cases} \tag{11}$$

We can see that the constraint condition of this dual problem is linear. The samples corresponding to $\alpha_i \neq 0$ are called support vectors. Use KKT condition to find the sample points $x_i$ landed on the optimal hyper sphere, These $x_i$ are satisfied with the constraint condition: $0 < \alpha_i < 1/vl$, $\sum_{i=1}^{l}\alpha_i = 1$. On the condition $\alpha_i < 1/vl$, quite a few of the sample points are outside of the hyper sphere, which are considered as abnormal sample points, while most of the sample points are within the sphere.

The decision function is as follows:

$$f(x) = R^2 - \|\varphi(x) - c\|^2 \tag{12}$$

When f (x) ≥ 0, φ(x) is within the hyper sphere, and is considered normal. Otherwise, φ(x) is outside the sphere, and is considered abnormal.

## FAULT DETECTION OF MARINE REFRIGERATED CONTAINERS BASED ON ONE-CLASS SVM

**Data sources:** The data come from Fault Analysis Experiment of the refrigerated containers. Among them, there are 1505 sets of normal condition and 426 sets of abnormal. Each set contains 14 variables and the data format is as follows:

$$X_i = \begin{bmatrix} T_{suc}^1 & T_{dis}^1 & T_{kin}^1 & T_{kout}^1 & T_{be}^1 & T_{af}^1 & T_{0in}^1 & T_{0out}^1 & P_{suc}^1 & P_{dis}^1 & P_{be}^1 & P_{af}^1 & T_{ain}^1 & T_{aout}^1 \\ T_{suc}^2 & T_{dis}^2 & T_{kin}^2 & T_{kout}^2 & T_{be}^2 & T_{af}^2 & T_{0in}^2 & T_{0out}^2 & P_{suc}^2 & P_{dis}^2 & P_{be}^2 & P_{af}^2 & T_{ain}^2 & T_{aout}^2 \\ . & . & . & . & . & . & . & . & . & . & . & . & . & . \\ T_{suc}^n & T_{dis}^n & T_{kin}^n & T_{\mathrm{kout}}^n & T_{be}^n & T_{af}^n & T_{0in}^n & T_{0out}^n & P_{suc}^n & P_{dis}^n & P_{be}^n & P_{af}^n & T_{ain}^n & T_{aout}^n \end{bmatrix}_{14\times n} \tag{13}$$

1000 sets data randomly selected from the 1505 normal samples are used to train the ONE-CLASS SVM model and the rest 505 sets are used to check the model. 426 sets of abnormal samples are used to check the abnormal recognition effect of the One-Class SVM model.

The samples are normalized before training and testing to eliminate the influence of different physical dimensions and the normalization method is as follows:

$$X^* = [X - \min(X)]/[\max(X) - \min(X)] \tag{14}$$

X* and X are the value after and before normalization respectively. Min(X) and max(X) are the minimum and maximum value of the samples.

**Determination of one-class svm kernel parameters:** Xu *et al*. (2008) found that RBF was the most suitable function for One-Class SVM, so in the experiment, RBF is introduced as the nonlinear mapping function to project the original data space to the feature space. Vapnik (2000) and Gary *et al*. (2002) considered that when RBF is used as SVM kernel function, its parameters can automatically determined and SVM has rapid training speed. But in our experiment, One-Class SVM fault detection rate varied widely when the parameters γ and v of RBF have different value. Therefore, we change the value of the parameters γ and v until the optimal parameters pack is found.

The parameters optimization algorithm of One-Class SVM's RBF kernel function is designed in this

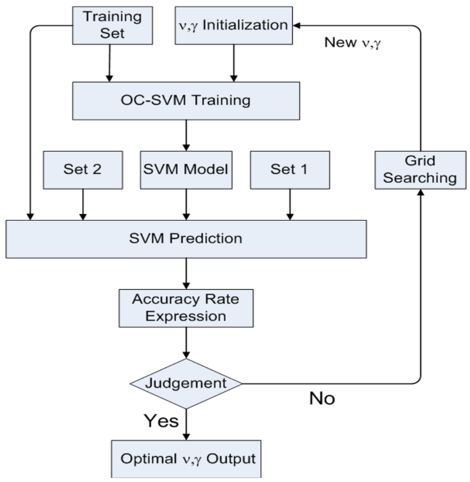


Fig. 1: One-Class SVM parameter optimization algorithm flow

study. The training effect of One-Class SVM is optimized through adjusting the kernel parameter γ and control parameter v. The algorithm flow chart is shown in Fig. 1.

Set 1 and Set 2 are the normal samples set and abnormal samples set.

Using this algorithm, optimal kernel parameters can be automatically determined within the setting range. The results of One-Class SVM experiment show that when γ ranges from 0.01 to 0.2 and v ranges from 0.01 to 0.1, One-Class SVM fault detection model has an accuracy rate of above 90% for training samples, a total detection rate of above 95% for testing samples, an acceptance rate of above 90% for normal samples in testing sets and a recognition rate of above 95% for abnormal samples in testing sets.

Therefore, after γ and v are primarily selected, we use the above algorithm to further optimize them with γ in the range of 0.01 to 0.2 and v in the range of 0.01 to 0.1. The experiment results are shown in Fig. 2.

Figure 2 shows that when γ ∈ [0.02, ,0.03] and v ∈ [0.01, 0.02] ,the accuracy rate for training samples and total detection rate for testing samples reach above 99% and 98%, respectively which are the highest and remain unchanged, the acceptance rate for normal samples can reach 100% and the recognition rate for abnormal samples reduces with the increase of

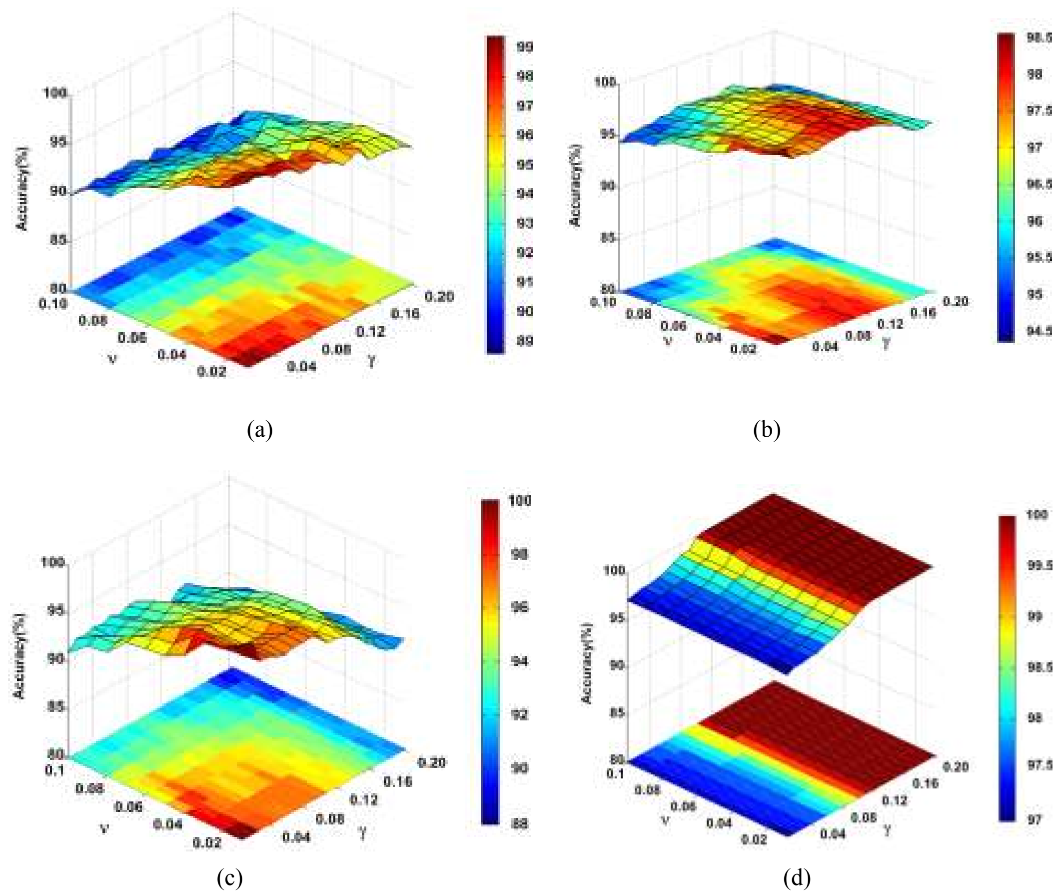


Fig. 2: One-class SVM fault detection rate; (a): Accuracy rate for training samples; (b) : Total detection rate for testing samples; (c): Acceptance rate for normal samples; (d): Recognition rate for abnormal samples v

γ and v has a highest value of 97.4% when γ takes 0.02 and v takes 0.01.

The research results also show that with the increase of γ, the training model has an increasing abnormal recognition rate, but a decreasing acceptance rate for normal samples. That is to say the detection rate for abnormal samples is inversely proportional to the acceptance rate for normal samples. To get a high abnormal detection rate must be at the expense of the normal recognition rate, i.e., increasing the possibility of misjudging normal condition for abnormal.

According to the experiment results and the feature of the above parameters, the selected RBF kernel function parameter γ is 0.02 and model control parameter v is 0.01. Meanwhile, the ONE-CLASS SVM model has a recognition rate for abnormal samples of 97.4% and an acceptance rate for normal samples of 100%, that is to say, the abnormal detection rate of the refrigerated container fault detection model is 97.4% with the missed alarm rate of 2.6%, and the normal recognition rate is 100% with false alarm rate of 0.

## CONCLUSION

The latest research results of statistical learning SVM is applied to the fault detection of refrigerated containers in the study, and the fault detection model based on One-Class SVM, a new SVM method is put forward to solve the problem of refrigerated containers that in actual operation there are only a large amount of normal samples, but few abnormal samples. The refrigerated container fault detection model based on One-Class SVM is established in the study and parameter optimization algorithm of RBF kernel function is designed. With the optimal model parameters, good experiment results are achieved that the recognition rate of abnormal samples reaches 97.4% and the acceptance rate of normal samples comes up to 100%, i.e., zero false alarm rate.

## REFERENCES

Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Disc., 2(2): 121-167.

Clarke, S.M., J.H. Griebsch and T.W. Simpson, 2005. Analysis of support vector regression for approximation of complex engineering analyses. J. Mech. Des., 127(6): 1077-1088.

Dong, J., A. Krzyżak and C.Y. Suen, 2005. An improved handwritten Chinese character recognition system using support vector machine. Pattern Recogn. Lett., 26(12): 1849-1856.

Fugate, M. and J.R. Gattiker, 2003. Computer intrusion detection with classification and anomaly detection, using SVMs. Int. J. Pattern Recogn. Artif. Intell., 17(3): 157-175.

Gary, G.L., N.H. Nelson and E.H. Norden, 2002. Application of the Hilbert-Huang transform to machine tool condition/health monitoring. AIP Conf. Proc., 615(1): 1711-1718.

Guyon, I., J. Weston, S. Barnhill and V. Vapnik, 2002. Gene selection for cancer classification using support vector machines. Mach. Learning, 46(1-3): 389-422.

Liu, J., J.P. Chen, S. Jiang and J. Cheng, 2003. Online LS-SVM for function estimation and classification. J. Univ. Sci. Technol., 10(5): 73-77.

Osuna, E. and R. Freund, 1997. Training support vector machines: An application to face detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 17-19 Jun, MA, USA, pp: 130-136.

Scholkopf, B., J.C. Platt, J.S. Taylor, A.J. Smola and R.C. Williamson, 2001. Estimating the support of a high-dimensional distribution. Neural Comput., 13(7): 1443-1471.

Schwenker, F., 2000. Hierarchical support vector machines for multi-class pattern recognition. Proceedings of the 4th International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies.

Sun, A., E.P. Lim and W.K. Ng, 2002. Web classification using support vector machine. Proceedings of the 4th International Workshop on Web Information and Data Management, pp: 96-99.

Tax, D.M.J. and R.P.W. Duin, 1999. Support vector domain description. Pattern Recogn. Lett., 20: 1191-1199.

Vapnik, V., 2000. SVM method of estimating density, conditional probability and conditional density. Proceeding of the IEEE International Symposium on Circuits and Systems, 5: 28-31.

Xu, T., Y. Luo and D.K. He, 2008. SMO training algorithm for hyper-sphere one-class SVM. Comput. Sci., 5(6): 178-180.

Zhong, Q.L. and Z.X. Cai, 2006. Sensor fault diagnosis based on one-class SVM. Comput. Eng. Appl., 19: 1-3.