Acoustic Emission Signal Classification based on Support Vector Machine

Yang Yu¹, Liang Zhou²
School of Information Science and Engineering, Shenyang University of Technology, China
No.111, Shenliao West Road, Economic & Technological Development Zone, Shenyang, P.R.CHINA
e-mail: yuy@sut.edu.cn¹, iceman0512@sina.com²

Abstract
Acoustic emission method has a major application in the detection of the oil storage tank damage. Therefore, classification of acoustic emission signals has great significance. A classification method based on support vector machines is proposed for its good generalization performance and less training data. Based on cross validation method, the genetic algorithm is compared with the grid search algorithm. The best parameters of the RBF kernel had obtained by using grid optimization method, and the classifier had built to achieve the identification and classification of acoustic emission signals. The simulation results show that support vector machine can effectively distinguish different acoustic emission signal and noise signal.

Keywords: acoustic emission signal; support vector machine; classification

1. Introduction
Acoustic emission detection of tank bottom damage is great importance on the oil storage and transportation, and the acoustic emission signal recognition research is the most important. Classification and recognition of acoustic emission signal include two aspects of quantitative and qualitative research. The quantitative analysis of acoustic emission source could reflect corrosion degree and qualitative identification refers to the nature of acoustic emission signals. The pattern recognition of acoustic emission signal includes amplitude identification, frequency identification, statistical pattern recognition, fuzzy diagnosis, artificial neural network approach and so on [1]. In practical applications, these methods often require researchers with rich background knowledge, as well as long-term acoustic emission site testing and data analysis experience, and this is its limitations. Artificial neural network has a good ability of self learning, but its minima local and promotion ability are narrow. Support vector machine method overcomes these deficiencies and has a good generalization performance of machine learning methods.

2. Research Method
2.1. Support Vector Machine
The feature of AE signals such as duration, amplitude, energy, frequency, rise time, ring count and impact number can be used as support vector machine classification dimension. To meet the needs of experiments, the classification may produce a higher number of dimensions. Kernel function of Support Vector Machine can handle the problem of high dimension. The kernel function is composed of two variables with nonlinear symmetric function. It can be the inner product of two vectors in a high dimensional feature space. So the computing complexity of optimal separating surface is the same, even if the transform space is in large dimensions. It is suitable for solving the problem of high dimensionality [2].

The support vector machine is first proposed by Vapnik. It bases on statistical principles and aims to achieve the structural risk minimization. The principle of support vector machine is the machine learning feature vector of train data, then it creates a classification model to classify. The main idea is to construct a hyperplane as a classification decision surface and make distance interval between each category to maximize Figure.1 [3]. H is the classification hyperplane. H1 is the minimum distance from the sample to the hyperplane. The mathematical
model of the classification hyperplane is as follows: Eq. 1 is the minimum distance of hyperplane and the classification model is Eq. 2.

\[
\min \phi(\omega, \xi) = \frac{1}{2} (\omega \cdot \omega) + C \sum_{i=1}^{l} \xi_i
\]

\[
y_i ((\omega \cdot x_i) + b) \geq 1 - \xi_i, i = 1,..,l \quad \xi_i \geq 0, i = 1,..,l
\]

\(x_i\) is the sample, \(y_i\) is the discriminant result, \(C\) is the punishment parameter, \(\xi_i\) is non-slack variable, \(b\) is the distance from origin to the decision, \(\omega\) is the hyperplane normal vector.

![Figure 1. The classification hyperplane](image)

The sample is inputted to the high-dimensional space, Eq. 2 is converted to the dual problem and to solve the discriminant function Eq. 3.

\[
y = \text{sgn} [ \sum_{x_i \in SV} y_i \alpha_i \phi(x_i \cdot x)] + b]
\]

\(\alpha_i\) is the Lagrange multiplier.

2.2. Kernel Function

\(K(x_i, x) = \langle \phi(x_i), \phi(x) \rangle\) is called the kernel function [4]. Which the kernel function has many types, such as linear kernel, polynomial kernel function, RBF kernel function, sigmoid kernel function:

\[
K(x_i, x_j) = x_i \cdot x_j
\]

\[
K(x_i, x_j) = [(x_i \cdot x_j) + c]^{q}
\]

\[
K(x_i, x_j) = \exp(- \|x_i - x_j\|^2 / (2\sigma^2))
\]

\[
K(x_i, x_j) = \tanh(b(x_i \cdot x_j) - c)
\]

The choice of kernel function determines the structure of the feature space. For the polynomial kernel function, the \(q\) value must be very great with the high dimensionality of the
feature space. It makes the calculation increasingl y, and some cases can’t get the correct results. Sigmoid kernel function has some limitations that the function \( b \) and \( c \) satisfy the Mercer conditions only in certain values. Polynomial kernel function has two controllable parameters \( q \) and \( c \). The RBF kernel function is only one parameter \( \sigma \). Therefore, parameters can reflect the complexity of model selection. From the number of parameters, the RBF kernel is a wise choice. It is a widely kernel function. By the choice of parameter, it can be applied to any distribution of the sample [4-6].

RBF kernel function parameter \( g \) is an artificial adjustable parameter, \( g = 1/2 \sigma^2 \). \( C \) is the penalty parameter. The choice of parameters determines the performance of the classifiers and different parameters will lead to the different classifiers and the effects. Therefore a good parameter selection method determines the classification results [6], [7]. The process of the RBF kernel function parameters selected depends on cross validation method. It can make the result more reliable by multiple operations.

2.3. Parameter Optimization Method

2.3.1. Cross Validation

Cross validation is a statistical analysis method used to verify the performance of the classifier. The basic original data grouping is with two parts. One part of the data is treated as the training set, while another part of the data is as a validation set. Training set is for classifier training and the validation set is for the model test. It can be used as the performance of the classifier [8]. The cross-validation methods are commonly used with three kinds:

Hold-Out Method : The original data are randomly divided into two groups, a training set, a validation set. They are used to train the classifier and validate the model. The accuracy of classification is used as this classifier performance. However, this method does not reflect the cross-ideological because the original data packet is random. The original data packet directly affects validation set level and classification accuracy. The results obtained by this method is not persuasive.

K-fold Cross Validation (K-CV): The original data are divided into \( K \) groups. Each subset of data is used as every validation set, and then the rest of the K-1 set of data is used as a subset of the training set. After processing the K model, its classification accuracy rate is seen as the average K-CV classifier performance[9]. The actual operation generally \( K \) is larger than 3. K-CV can effectively reduce the over-learning and less learning. The final result is more persuasive.

Leave-One-Out Cross Validation (LOO-CV): If the original data have \( N \) samples, LOO-CV has \( N \) validation sets. The remaining \( N-1 \) samples are used as the training set. LOO-CV method will get \( N \) model. The average classification accuracy of \( N \) models can be used as the performance of the classifier (LOO-CV). But the drawback of the LOO-CV is for its high computational cost. If the number of data samples is excessive, LOO-CV implementation will be difficult and impossible. So this paper uses the K-CV method to verify the classifier performance.

Based on the idea of cross validation, two parameter optimization methods will be used. The parameter optimization methods are commonly used with two kinds:

2.3.2. Grid Optimization

The grid search method is as follows: At the first, \( C \) and \( g \) are taken to different \( M \) and \( N \) values and divided into \( N \times M \) combinations (\( C \) and \( g \)). Then a group of \( C \) and \( g \) is selected, such as \( g = 2^{10} \sim 2^{15} \), the search step is set as -1; \( C = 2^{-10} \sim 2^{15} \), the search step is 1; A two-dimensional grid is constituted on the coordinate system by \( C \) and \( g \). It corresponds to grid on each set of \( C \), \( g \) value. In accordance with the method described above, it calculated the prediction accuracy rate. Finally, contour lines can be draw based on each group \( C \) and \( g \) values correspond to the accuracy. According to the contour map, the best \( C \) and \( g \) value can be determined. If the accuracy can not meet the requirement, a search area can be selected based on the existing contour map, and then reduced the search step size for better search. The advantage of the grid search method is to search for \( C \) and \( g \) values at the same time. Ultimately the best parameters have been obtained to achieve the best classification accuracy of the discriminant function [9].
2.3.3. Genetic Algorithms Optimization.

The genetic algorithm is a directly search method independent on the specific issues. The evolution of populations is achieved by natural selection, exchange and variation utilizing the biological genetics view and combining with the ideas of fittest survival and random information exchange. In the optimization process, the GA randomly generates more than one starting point. At the same time, it starts to search in the solution space and the fitness function can guide the search direction. Firstly it selects initial population randomly and calculates the individual fitness function value. If the best individual in the population corresponding to the fitness function value is large enough or algorithm has been run continuously multi-generational, and the best individual fitness is no significant change, the best kernel function parameters $C$ and $g$ can be obtained [10].

For optimal selection of SVM parameters, the international community hasn’t recognized the uniform. Currently common method is to select $C$ and $g$ in a certain value range, then take the training set as the original interest data set under the group of $C$ and $g$, use cross-validation method to get its classification accuracy at last. Finally it achieved the training set to verify the group with the highest classification accuracy rate of $C$ and $g$. An accurate rate may correspond to multiple sets of parameters $C$ and $g$. So it needs to select the smallest parameter $C$ as the best parameters. If there are many groups of $g$ corresponded to the smallest $C$, the first group of $C$ and $g$ is selected as the best parameters. If $C$ is too high, it will lead the over-learning state, namely, classification accuracy of the training set is high whereas test set classification accuracy rate is very low (the generalization ability of classifiers reduced). In all pairs of $C$ and $g$ with the highest validation classification accuracy, the small penalty parameter $C$ is a better choice of object [11],[12].

3. Results and Analysis

The acoustic emission signal is a transient signal. The mathematical model is Eq. 8:

$$s(t) = A_0 + \sum_{i=1}^{l} A_i \exp[-a_i(t - t_i)] \cdot \sin[\omega(t - t_i) + \varphi_i]$$  \hspace{1cm} (8)

$s(t)$ is the acoustic emission, $A_i$ is the amplitude of the signal, $a_i$ is the signal attenuation coefficient, $t_i$ is the time of arrival for the signal, $\omega$ is the frequency of the signal. It uses mathematical models to simulate a continuous acoustic emission signal.

Pencil lead break signal is used to simulate burst acoustic emission signal. The chaotic noise signal is simulated by using sandpaper to rub surface of the plate. Experiment collects the pencil lead break signal and friction signal for 100 times each. Support vector machine can distinguish the noise signal from the burst acoustic emission signals and the continuous acoustic emission signal.

![Figure 2. C and g parameters of GA optimization](image1)

![Figure 3. Classification results of GA optimization](image2)
Acoustic Emission Signal Classification based on Support Vector Machine (Yang Yu)

4. Conclusion

In this paper, an optimum classifier is obtained by researching support vector machine and analyzing key parameters of the kernel function. Using cross-validation method, the average of multiple compute makes accuracy more credible. The classification accuracy rate of grid search parameters is higher than the GA algorithm by this method. So the grid optimization method is more suitable for the classification of the acoustic emission signals. The simulation results show that the support vector machine can distinguish noise signals and the different kinds of acoustic emission signals with 98.88% classification accuracy rate. So it is worth promoting for tank bottom damage acoustic emission signals classification.

Acknowledgements

This research was financially supported by the Creative Team Project Foundation of Education Department of Liaoning Province of China (Grant No.LT20100082).

References


