Leak Detection in Transport Pipelines Using Enhanced Independent Component Analysis and Support Vector Machines*

Zhengwei Zhang, Hao Ye, Guizeng Wang, and Jie Yang

Department of Automation, Tsinghua University, Beijing 100084, P.R. China
zhangzw02@mails.tsinghua.edu.cn

Abstract. Independent component analysis (ICA) is a feature extraction technique for blind source separation. Enhanced independent component analysis (EICA), which has enhanced generalization performance, operates in a reduced principal component analysis (PCA) space. SVM is a powerful supervised learning algorithm, which is rooted in statistical learning theory. SVM has demonstrated high generalization capabilities in many pattern recognition problems. In this paper, we integrate EICA with SVM and apply this new method to the leak detection problem in oil pipelines. In features extraction, EICA produces EICA features of the original pressure images. In classification, SVM classified the EICA features as leak or non-leak. The test results based on real data indicate that the method can detect many leak faults from a pressure curve, and reduce the ratio of false and missing alarm than conventional methods.

1 Introduction

With the development of the transport pipeline industry, leak detection in transport pipelines has drawn intensive attention. For all pipeline leak detection techniques, the negative pressure wave based approach has been widely adopted in real applications.

The basic idea of the negative pressure wave based methods is that, when leak occurs, there is a negative pressure wave propagating from the leak point toward the upstream and downstream ends and so that we can detect leaks by detecting the pressure changes at the both ends [1].

Many approaches have been introduced for leak detection based on negative pressure wave, such as Wavelet Transform [2], Pattern Recognition [3] etc.. However, the performance of these methods is often bound by high false alarm/miss detection rates.

Recent advances in statistical learning theory, especially, the introduction of support vector machines (SVM) has made it possible to get high accuracies for the pattern recognition problem [4]. Enhanced independent component analysis (EICA) is also a very powerful tool for blind signal separation with better generalization performance than Independent component analysis (ICA) [5]. In this paper, we propose a novel Leak Detection method based on EICA/SVM method which integrates EICA with SVM, and apply it to the leak detection problem. The improved performance has been demonstrated by tests based on real data.

---

* Supported by the National Natural Science Fund of China (60274015) and the 863 Program of China.

L. Wang, K. Chen, and Y.S. Ong (Eds.): ICNC 2005, LNCS 3611, pp. 95 – 100, 2005.

2 Theoretical Background

2.1 Enhanced Independent Component Analysis

ICA, which expands on principal component analysis (PCA) as it considers higher order statistics, is originally developed for blind source separation whose goal is to reduce statistical dependencies and derives a sparse and independent but unknown source signals from their linear mixtures without knowing the mixing coefficients.

Let us assume a linear mixture model

$$\mathrm{X} = AS \tag{1}$$

where X denote the linear mixtures, S denote the original source signals whose components are independent and unknown, and A is unknown. The goal of ICA is to estimate the matrix W in the reconstruction model

$$Y = W\mathrm{X} \tag{2}$$

A large amount of algorithms have been developed for performing ICA. One of the best methods is the fixed-point-FastICA algorithm proposed by Hyvärinen [6]. FastICA is a method by which the independent components are extracted one after another by using Kurtosis. This method has high-speed convergence.

In order to improve the generalization performance of ICA, Liu et al have proposed a novel EICA method [5]. EICA, whose enhanced retrieval performance and reduced computational complexity are achieved by means of generalization analysis, operates in a reduced PCA space.

Let C denote the covariance matrix of X. We can get the orthogonal eigenvector matrix Ψ and a diagonal eigenvalue matrix Λ with diagonal elements in decreasing order by using PCA. Now let P be a matrix whose column vectors are the first n leading eigenvectors of C,

$$P = [\psi_1, \psi_2, \ldots, \psi_n] \tag{3}$$

where P is the loading matrix, n is determined by balancing two criterions of the PCA procedure: for image feature representation and for ICA generalization [5].

The new random vector Z in this reduced PCA space is defined as

$$Z = P^T \mathrm{X} \tag{4}$$

We call the ICA method implemented in this reduced PCA space the EICA method. A more detailed description of the EICA algorithm can be found in [5].

2.2 Support Vector Machines

SVM was originally introduced by Vapnik [4], which is a novel type of learning machine based on Statistical Learning Theory. The SVM aims at minimizing an upper bound of the generalization error through maximizing the margin between the separating hyperplane and the data. Such a scheme is known to be associated with structural risk minimization [4].

Generally, in a SVM classifier, the discriminant function has the following form:

$$f(x) = sgn(\sum_{i=1}^{M} \alpha_i^* y_i K(x_i, x) + b^*) \tag{5}$$

Where the parameters are obtained by maximizing the objective function:

$$\begin{aligned} Q(\alpha) &= \sum_{i=1}^{M} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{M} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ subject \quad to \quad & \sum_{i=1}^{M} \alpha_i y_i = 0 \quad 0 \le \alpha_i \le C \quad i = 1, \ldots, M \end{aligned} \tag{6}$$

By solving the above quadratic programming problem, SVM tries to maximize the margin between data points in the two classes and minimize the training errors simultaneously. For the nonlinear case, the training patterns are mapped onto a high-dimensional space using a kernel function. In this space the decision boundary is linear. The most commonly used kernel functions are polynomials, exponential and sigmoidal functions. Here we choose radial basis function (RBF) kernel function as

$$K(x_i, x_j) = \exp(-\left\| x_i - x_j \right\|^2 / N\sigma^2) \tag{7}$$

where N is the number of independent image basis which we extract with EICA.

3 Leak Detection Based on EICA/SVM

In [7], we treated the pressure curve of fixed length as an image and proposed an Eigencurves method based on EICA to detect leaks in oil pipelines. To achieve reduced computational complexity and better classification robustness, in this paper we presents a novel EICA/SVM based method which uses SVM to determine the appropriate class: leak or non-leak. The method includes two steps: (i) EICA based feature extraction; (ii) SVM based classification of EICA features obtained from step (i).

3.1 EICA Based Feature Extraction

Let $\chi \in \Re^{N^2}$ be a vector representing a pressure image with $N \times N$ pixels. The vector is formed by concatenating the rows or the columns of the image. Let the training set of pressure images be $\mathrm{X} = \{\chi_1, \chi_2, \cdots, \chi_M\}$, where M is the number of images in the training set.

Since we assume the pressure images are linear combinations of independent image bases in the ICA algorithm, we do not lose information by replacing the original pressure images with their eigenvectors that are linear combinations of the original pressure images [8]. Following this scheme, we firstly compute the loading matrix P [Eq. (3)] of the X, then apply the EICA algorithm as follows:

$$\begin{aligned} WP^T &= Y \\ so \quad P^T &= W^{-1} Y \end{aligned} \tag{8}$$

where each rows of Y represents an independent image basis [8].

The principal component representation of the set of zero mean images in X^T based on P is defined as $R = \mathrm{X}^T P$. A minimum squared error approximation of X^T is obtained by ${\mathrm{X}^T}_{rec} = RP^T$ [8]. From P, a minimum squared error approximation of X^T is obtained by

$$ {\mathrm{X}^T}_{rec} = \mathrm{X}^T P P^T \tag{9} $$

From Eq. (8) and (9), we can get,

$$ {\mathrm{X}^T}_{rec} = \mathrm{X}^T P W^{-1} Y = CY \tag{10} $$

where PW^{-1} is obtained during the EICA training procedure and each row of C is the EICA feature of corresponding training image. We take PW^{-1} as extraction matrix. For a test image $f \in \Re^{N^2 \times 1}$, we can extract EICA features by project f on PW^{-1}:

$$ c = f^T P W^{-1} \tag{11} $$

3.2 SVM Based Classifier Training

Since we have obtained EICA features matrix $C = [c_1^T, c_2^T, \cdots, c_M^T,]^T$ [Eq. (10)], we build the SVM training set $\{c_i, d_i\}_{i=1}^M$, where $d_i = \{\pm 1\}$ is the class type of EICA feature c_i. (+1means non-leak, -1means leak)

The support vectors and other parameters in the decision function $f(x)$ [Eq. (5)] are determined through numerical optimization during the classifier training step.

3.3 Online Detection

To realize online detection, this system detects leak by sampling the overlapping M (M=2000) width windows located on the pressure curve which is updated real-timely and classifying them using a trained SVM to determine the appropriate class: leak or non-leak. The window moves on the pressure curve with step of 1000 pressure values and new samples are extracted real-timely.

Applying the thought above, for an unknown pressure curve image f^T, we can get its EICA features c [Eq. (11)]. Then we can decide whether there is any leak in Transport Pipelines by using the nonlinear decision function [Eq. (5)].

4 Experimental Results

4.1 Data Preparation and Sample Set

In this paper, the pressure curve is formed by ordinal connecting 2000 continually sampled pressure data. The sample time is 60 ms. we fragment the original pressure curve image into several 30×30 images and treat them as training and test images set.

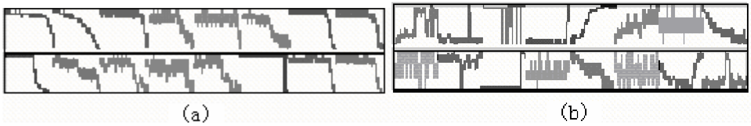

Fig. 1. (a) The negative pressure wave samples. (b) The normal condition samples.

Fig.1 shows some sample images we get by fragmenting one-hour pressure curve obtained from a real oil pipeline.

In this paper, we intercept 5,380 images as sample set in which 1,600 images contain negative pressure wave. We pick 1,500 images (500 images contain negative pressure wave) as training samples and the remnant are used as test images set.

4.2 EICA/SVM Based Negative Pressure Wave Detection

From the training images set, we extract 161 independent basis images with EICA. Fig.2 shows the first 45 learned independent basis images of them.

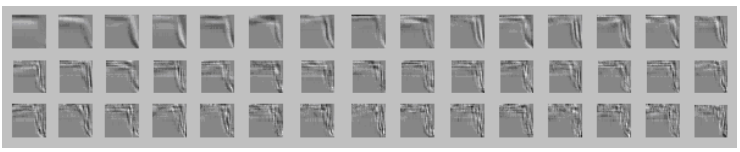

Fig. 2. First 45 learned independent basis images

We can see these independent basis images express more local feature, which suggests that they can lead to more precise representations.

Using the EICA features [Eq. (10)], the SVM is trained for classification. The SVM classifier uses the RBF kernel with $\sigma^2 = 8$ and the upper bound $C = 500$ to obtain a perfect training error. The training result is that there are 109 supports vectors which are about 7.27% of the total number of training images.

4.3 Results

In Table 1 we compared the EICA/SVM training and testing results with direct SVM. It can be seen that the EICA/SVM training achieves fewer support vectors. This suggests EICA/SVM have better generalization capacity. From this table we also see that the EICA/SVM scheme can reduce the false alarm/miss detection rates effectively.

Table 1. Test results

Approach	Support Vectors	Miss Detections	False Detections	Correct %
EICA/SVM	109	72	35	97.24
Direct SVM	341	187	167	91.24

5 Conclusion

In this paper, a new EICA/SVM method to detect the negative pressure wave is proposed. The experiment results show that the EICA/SVM method performs better than direct SVM method. The reason lies in the fact that EICA can explore edge information in the image data. By using EICA feature instead of the original image data, SVM reduced the number of support vectors in training and get lower detection errors.

References

1. Turner, N. C.: Hardware and software techniques for pipeline integrity and leak detection monitoring, Proceedings of Offshore Europe 91, Aberdeen, Scotland (1991)
2. Ye, H., Wang, G. Z., Fang, C. Z.: Application of Wavelet Transform to Leak Detection and Location in Transport Pipelines [J]. Engineering Simulation, Vol.13 (1995)1025-1032
3. Jin, S.J., Wang, L.N., Li, J.: Instantaneous Negative Wave Pattern Recognition Method in Leak Detection of Crude Petroleum Transported Pipeline, Journal of Electronic Measurement and Instrument, Vol.12 (1998) 59-64. (In Chinese)
4. Vapnik, V.: The nature of statistical learning theory, Springer, New York (1995)
5. Liu, C.J.: Enhanced independent component analysis and its application to content based face image retrieval, IEEE Trans. Systems, Man, and Cybernetics-part B: Cybernetics, vol. 34, no. 2 (2004)
6. Hyvärinen, A.: Fast and robust fixed-point algorithm for independent component analysis, IEEE Trans. Neural Networks, vol. 10 (1999) 626-634
7. Zhang, Z. W., Ye, H., Hu, R.: Application of Enhanced Independent Component Analysis to Leak Detection in Transport Pipelines, ISNN (2) (2004) 561-566
8. Bartlett, M.S., Lades, H.M., Sejnowski, T.J.: Independent component representation for face recognition [A], Proceedings SPIE Conference on Human Vision and Electronic Imaging Ⅲ [C], San Jose, CA, USA (1998) 528-539