

An SVM-Based Recognition Method for Safety Monitoring Signals of Oil and Gas Pipeline

Yang Chen^a, Jianhui Zhao^{*}, Fan Li^b

School of Instrumentation Science and Optoelectronics Engineering, Beihang University, Beijing 100191, China

*Corresponding author e-mail: zhaojianhui@buaa.edu.cn,

^achenyang1617@buaa.edu.cn, ^blifan@buaa.edu.cn

Abstract. An SVM-based recognition method for the safety of oil and gas pipeline was proposed due to limitation of the traditional learning methods based on empirical risk minimization. The vibration signals along the pipelines are obtained with the distributed optical fiber vibration sensor on the basis of Mach-Zehnder optical fiber interferometer theory. The wavelet packet threshold denoising is used to preprocess the signal. Then the eigenvectors of vibration signals were extracted through the energy-pattern method based on wavelet packet decomposition. At last the vibration signals were recognized by support vector machine (SVM) through the eigenvectors with a view to detecting whether abnormal events happened along the pipelines.

1. Introduction

In recent years, pipeline leaking accidents occur frequently because of human damaging activities and construction around pipelines. The current research on safety monitoring of buried oil and gas pipeline includes the pressure gradient method, the negative pressure wave method, quality/flow balance method, ultrasonic detection and etc [1]. However, these leaking detection methods cannot alarm before pipeline is damaged. What is more important, the accuracy of alarming and location cannot meet the oil and gas pipeline safety requirement. The advent of fiber sensor, with advantages of simple structure, low price, long-distance detection and accurate positioning [2], has opened a plethora of safety monitoring. A distributed optical fiber alarming system for the safety of oil and gas pipeline has been developed. By capturing and analyzing along the pipeline vibration signal (leakage or external intrusion events all can produce corresponding vibration). Besides, this system make a reservation to leak position and implement early warning monitoring. It is one of the core contents of this system to accurately judge whether abnormal events occur along the pipeline.

Traditional learning methods are mainly based on empirical risk minimization (ERM) principle [3]. As the training data tends to infinity, the empirical risk converges to the actual risk. Therefore, the empirical risk minimization principle implicitly USES the hypothesis condition that the training sample tends to infinity [4]. But in practical engineering, the number of training samples available is always limited, so the traditional algorithm encountered many difficulties, such as small sample problems, over fitting problems, high dimensional problems and learning machine structure problems and so on. Support vector machine (SVM) is a learning method based on statistical theory. Compared with the traditional learning method, it is based on the principle of structural risk minimization, thus it



can make a reasonable compromise between the empirical risk and the model complexity so as to obtain higher promotion ability.

Firstly, the signal is preprocessed based on wavelet packet threshold denoising, and the "energy-mode" method of wavelet packet decomposition is used to extract the eigenvector of the vibration signal, which can extract the vibration signal characteristics rapidly and efficiently. It can overcome the shortcomings of traditional learning methods and realize the fast and accurate identification of abnormal events along the pipeline.

2. Model for pipeline safety monitoring

2.1. System detection principle

We develop a model for oil and gas pipeline safety monitoring system, as shown in Figure 1. The fiber optic cable is laid along the pipeline in parallel, using three single fiber membranes which compose the distributed optical fiber vibration signal sensor based on Mach-Zehnder fiber interferometer principle. Two of the optical fibers are sensor fiber, and the third fiber is used for signal transmission [6]. Two light waves in optical fiber come together and format the interference signal which is transmitted to the photodiode. The optical signal is converted to electrical signal, then the signal are processed through the amplification and filter circuit, after which A/D acquisition convert it to the computer for further signal processing and analysis.

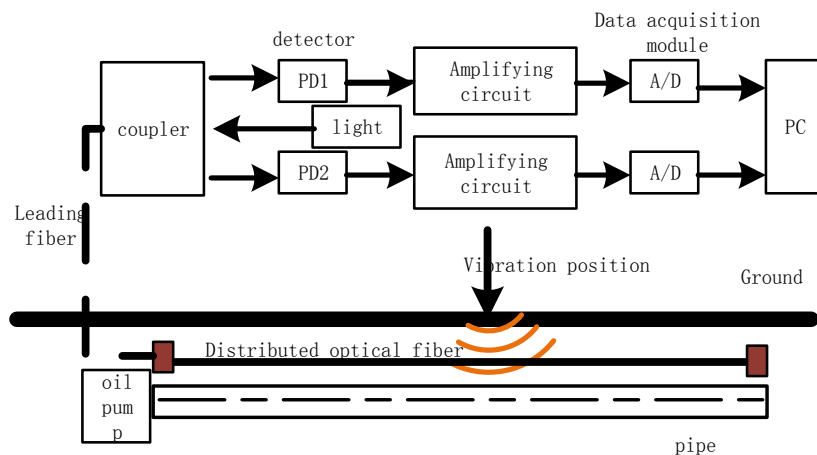


Fig. 1 Signal collection mechanism

When the fiber is affected by the vibration signals along the pipeline, since two test fibers are arranged in different positions in the cable, and these two test fibers will produce different strains. Therefore, the coherent light waves in the two test fibers will produce different phase changes respectively [7].

After intervention at the end of the sensing fiber optic cable, the light intensity is:

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos[\Delta s(t) + \Delta\varphi] \quad (1)$$

Where: $\Delta s(t)$ is the difference between two beam interference light wave phase modulation; $\Delta\varphi$ represents the difference between the initial phase; I_1 and I_2 are light intensity of two coherent waves. Suppose I_0 is the total input light intensity of two test fibers, there are

$$I(t) = I_0 \{1 + \alpha \cos[\Delta s(t) + \Delta \varphi]\} \quad (2)$$

α represents the mixing efficiency of two coherent light waves

If only the light intensity is considered, equation (2) can be simplified as:

$$I(t) = I_0 \alpha \cos[\Delta s(t) + \Delta \varphi] \quad (3)$$

The photoelectric detector converts the light intensity signal into the current signal, and the amount of light current is

$$I(t) = KI_0 \alpha \cos[\Delta s(t) + \Delta \varphi] \quad (4)$$

Where K is the photoelectric conversion coefficient. $\Delta s(t)$ is a variable and the test signal is a function of $\Delta s(t)$. By detecting the change of interference light signal in real time, the vibration signals along the pipeline can be monitored, so as to realize real-time monitoring and warning of pipeline leakage.

2.2. Signal Preprocessing

Wavelet packet analysis can provide a detailed analysis method for signal, the band is divided into multiple layers, the high frequency part is further decomposed, and the frequency band can be selected according to the characteristics of the analysis signal, so as to match the signal spectrum, thus improving the ability of processing signal.

The characteristics of wavelet packet transform can decompose the original signal into a series of approximation components and detail components, and large amplitude in general is given priority to with the useful signal, the amplitude of the smaller coefficient is likely to be noise. Therefore, the threshold value method can be used to keep the signal coefficient and reduce the most noise coefficient to zero. After the reconstruction of the wavelet packet, the smooth signal can be obtained. Follow these steps:

1. Select a small wave base and determine the number of layers of a wavelet decomposition and then decompose the signal N layer wavelet packet.
2. Compute the best tree (that is, determine the best wavelet packet) for a given entropy standard and calculate the best wavelet packet decomposition tree.
3. Select an appropriate threshold for threshold quantization of the high frequency coefficients at each decomposition scale. In order to obtain the optimal result to meet the specific analysis and information evaluation criteria, the threshold should be determined through repeated experiments.
4. Wavelet packet reconstruction according to the N layer of wavelet packet decomposition coefficient of low frequency and quantitative processing coefficient.

2.3. Vibration signal feature extraction.

The characteristic of vibration signal is extracted by "energy-mode" method based on wavelet packet decomposition. Suppose the signal sampling frequency is $2f$, if the signal is decomposed by j layers wavelet packet, then 2^j wideband bands can be formed, and the bandwidth of each interval is $\frac{f}{2^j}$. After the wavelet packet is decomposed, j layers of wavelet packet coefficient $C_{j,k}^m$ can be obtained, among which k is the sequence number of the decomposition node of the layer, and m is the location identification of the wavelet packet space.

The energy of the signal $x(t)$ in the time domain is

$$\|x(t)\|^2 = \int_{-\infty}^{+\infty} |x(t)|^2 dt \quad (5)$$

Equation (5) and the wavelet transform coefficient of $x(t)$ are associated with the Parseval energy integral equation. We can get

$$E_{j,k} = \sum_m |C_{j,k}^m|^2 \quad (6)$$

As can be seen from formula (6), the wavelet transform coefficient $C_{j,k}^m$ has energy dimension, which can be used to analyze the energy of signal.

The signal energy is selected as the characteristic parameter of vibration signal, and the feature vector extraction steps based on wavelet packet decomposition are wavelet packet decomposition for vibration signals; select multiple frequency bands with the most sensitive signal energy, and find the energy of each band and normalize it, namely

$$T_{j,k} = \sum_m |C_{j,k}^m|^2 \quad (7)$$

$$T'_{j,k} = \frac{T_{j,k}}{\sum_n T_{j,k}} \quad (8)$$

The above normalized energy is used as the characteristic vector of the vibration signal.

$$T = [T'_1, T'_2, \dots, T'_n] \quad (9)$$

2.4. An event classification method based on support vector machine

Support vector machine (SVM) is a learning method developed on the basis of statistical theory. The basic idea of SVM is that it will transform the input space to a high-dimensional feature space by nonlinear transformation. We can calculate the optimal separating hyper plane in that feature space and makes linear inseparable data in the original input space into a linear separable [9], and using minimum error theory structure instead of minimum expected error, effectively avoid over fitting problem in the neural network theory.

The two-category pattern recognition issue of support vector machine (SVM) is the basis of solving multi-classification problem. In the case of dichotomous linear separable problem, the linear separable sample is assumed to be $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l), x \in R^n, y \in \{-1, 1\}$. The goal of sample learning is to construct a decision function that classifies the test data as correctly as possible. The classification equation of the dichotomous linear problem is

$$y((w \cdot x_i) + b) - 1 \geq 0 \quad i = 1, 2, \dots, l \quad (10)$$

There are two conditions for constructing optimal classification surface, namely empirical risk minimization principle and the minimum of confidence range. The empirical risk minimization principle satisfies formula (10). The minimum conditions of confidence range can be converted into:

$$\Phi(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 = \frac{1}{2} (\mathbf{w} \cdot \mathbf{w}) \quad (11)$$

Therefore, in the case of the linear, the optimal classifications are actually the problem of solving the quadratic constraint optimization down here, which is

$$\begin{aligned} \min \Phi(\mathbf{w}) &= \frac{1}{2} \|\mathbf{w}\|^2 = \frac{1}{2} (\mathbf{w} \cdot \mathbf{w}) \\ \text{s.t. } y_i(\mathbf{w} \cdot \mathbf{x}_i + \mathbf{b}) - 1 &\geq 0 \quad i=1,2,\dots,l \end{aligned} \quad (12)$$

The constraint problem is transformed into a Lagrange function solution, which can be transformed into the following by the duality principle

$$\begin{aligned} \max \mathbf{w}(\alpha) &= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \\ \text{s.t. } \sum_{i=1}^l y_i \alpha_i &= 0 \quad \alpha_i \geq 0, i=1,2,\dots,l \end{aligned} \quad (13)$$

Formula (13) is a quadratic programming problem, and the optimal solution α^* can be obtained by solving it. Then we can find \mathbf{w}^* 、 \mathbf{b}^* and construct the hyper plane. The corresponding classification discriminant function is

$$f(\mathbf{x}) = \text{sgn}\left\{\sum_{i=1}^l \alpha_i^* y_i (\mathbf{x}_i \cdot \mathbf{x}) + \mathbf{b}^*\right\} \quad (14)$$

The problem of linear inseparability can be solved by introducing a relaxation term $\xi_i \geq 0$ in its classification equation.

When the sample set is nonlinear separable. The input sample is often mapped to a higher dimensional feature space, and the optimal classification surface is constructed in this space and solved. The support vector machine has transformed the classification problem into a constrained optimization problem, namely quadratic programming problem. Many algorithms in optimization theory need to take advantage of the entire Hessian matrix, so these algorithms are only applicable to small training sets. When the size of the sample set is large, this approach increases the system's consumption dramatically, so these methods are not suitable for processing large-scale data. In this paper, sequence minimum optimization (SMO) [11] is adopted to solve the above problems.

Abnormal event identification along pipeline is a typical multi-classification problem. To solve multi-classification by support vector machine (SVM) [12], "one to one" method to construct multivariate classifier, a single SVM training scale is small, training data is balanced, and it is easy to expand at the same time. So we adopt the "one to one" strategy to solve the problem of the multi-classification of the abnormal events along the pipeline.

3. Experiment and data feature extraction

3.1. Experiment

We carry out relevant simulation experiments in the laboratory, respectively manufacture hammer hit laying pipeline, human walking and field noise 3 cases. The pipeline is laid in the same direction as the optical cable, and the optical cable is directly above the pipeline, as shown in figure 2.



Fig. 2. Experiment Example of a figure caption.

3.2. Signal feature extration

In the "energy-mode" method, db3 wavelet function is used to extract the vibration signals caused by the above three conditions. The one-dimensional vector composed of normalized energy of eight sensitive bands is obtained. In the field experiment, the system collects vibration signal and its characteristic vector as shown in figure 3. The vibration signal characteristic results show that the characteristic vector of the vibration signal can be obtained by this method.

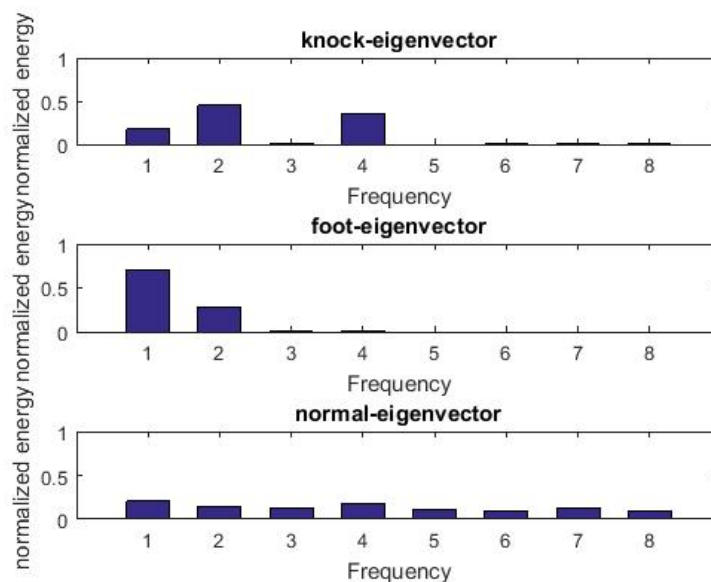


Fig. 3 The eigenvectors of the signal in three cases

3.3. *Events area classified using support vector machines*

100 groups are randomly selected from a large number of data of each action in the field experiment for the learning of SVM. In addition, 30 groups of data of each action are randomly selected for testing the support vector machine that completed the learning. It is trained by "one-to-one" algorithm. .

30 groups of test data were identified by the support vector machine that had completed the sample learning. After the training, three samples were incorrectly identified, and the recognition accuracy is 90%.

4. Summary

The recognition method based on support vector machine is adopted for the typical multi-classification issue of abnormal event warnings along the pipeline. In this paper, the "energy-mode" method based on wavelet packet decomposition is used to extract the characteristic of detection signal. Then the support vector machine is used to identify the abnormal events along the pipeline and solve the multi-classification problem by using the "one-to-one" method. The above methods are verified by field experiment data, and the results show that the recognition method based on support vector machine has good real-time performance and high accuracy. Therefore, this method can be used to identify the abnormal events along the pipeline, so as to protect the safety of pipeline. In the future work, more types of abnormal event identification work needs to be studied and improved through further field experiments.

References

- [1] Jie Yang, G.Z. Wang, "Leak Detection and Location Methods for Gas Transport Pipeline", Control and Instruments in Chemical Industry, April 2004, Vol. 31, No.3, pp. 1 - 5.
- [2] M.C. Maki, J.K. Weese, "Fiber Optic Fence Sensor", IEEE AESS System Magazine, February 2004, pp. 8 - 13.
- [3] Li Wang, Y.H. Wang, X.L. Xiao, Hu Yan, G.S. Shi, Q.R. Wang, "A Fiber-sensor-based Long-distance Safety Monitoring System for Buried Oil Pipeline", Proceedings of 2008 IEEE International Conference on Networking, Sensing and Control, Sanya, China, April 6-8 2008, pp. 451 - 456
- [4] Cortes C, Vapnik V. Support-vector networks [J]. Machine Learning, 1995, 20: 273-297.
- [5] Hsu C W, Lin C J. A comparison of methods for multiclass support vector machines [J]. IEEE Transactions on Neural Networks, 2002, 13 (2): 415-425.
- [6] Cortes C, Vapnik V. Support-vector networks [J]. Machine Learning, 1995, 20: 273-297.
- [7] Deng Naiyang, Tian Yingjie. A New Method in Data Mining: Support Vector Machine [M]. Beijing: Science Press, 2004 (in Chinese).
- [8] Qu Zhigang. Study on the Leakage Detection and Pre-Warning Techniques Based on the Distributed Optical Fiber for the Long-Distance Oil and Gas Pipelines [D]. Tianjin: School of Precision Instrument and Opto-Electronics Engineering, Tianjin University, 2007 (in Chinese).