Feature Extraction and Identification of Pipeline Intrusion Based on Phase-Sensitive Optical Time Domain Reflectometer

Zhanfeng Zhao, Duo Liu(✉), Longwei Wang, and Shujun Liu

Harbin Institute of Technology (Weihai), Weihai, China
1398108921@qq.com

Abstract. Since fiber distributed vibration sensing (DVS) system based on phase-sensitive optical time domain reflectometer (Φ-OTDR) has the characteristics of identifying intrusion signals, wide monitoring range and high system sensitivity, correct identification of intrusion types by the system is an important issue to promote the engineering of this technology. In this paper, based on the intrusion signal of Φ-OTDR system, a multi-dimensional feature extraction and selection method is proposed. The polynomial least squares method is used to remove the trend term from the vibration signal, and the wavelet threshold denoising method is used to reduce the noise interference. The short-time analysis in the time domain and the wavelet analysis in the wavelet domain are combined to extract the multi-dimensional characteristics of the signal. The feature selection is based on the QUICKREDUCT algorithm. The experimental results show that the feature vector obtained by this method is relatively complete, and it is less affected by the environment, and the recognition rate is higher, reaching over 92%.

Keywords: Φ-OTDR · Short-term analysis · Wavelet analysis · Feature extraction · Identification

1 Introduction

As fiber distributed vibration sensing (DVS) system based on phase-sensitive optical time domain reflectometer (Φ-OTDR) has the advantages of low cost, long service life, corrosion resistance, good electrical insulation and strong concealment, it can be applied to long-distance perimeter security, pipeline pre-warning, and quantitative vibration measurement. In the fiber pre-warning system, the classification and identification of external intrusion signals is very important. If a false alarm occurs, it will not only cause waste of manpower and material resources, but also may delay the processing time and even endanger the safety of life and property. Therefore, how to accurately identify the types of intrusion events has always been the key issue in the research of optical fiber pre-warning systems, and has received extensive attention from researchers [1].

In previous studies, Mahmud et al. generated a dynamic threshold based on the measured signal. The number of times a signal of a certain duration passes through the threshold and the length of time exceeding the threshold are used as signal

M. Jia et al. (Eds.): WiSATS 2019, LNICST 280, pp. 665–675, 2019.
https://doi.org/10.1007/978-3-030-19153-5_65

characteristics [2]. However, this method only extracts the time domain features of the signal and cannot fully represent the signal characteristics [3]. Xu used convolutional neural networks to identify and classify time-frequency analysis patterns of different intrusion methods, but the number of training samples required by convolutional neural networks is large, and the time-frequency analysis map is susceptible to noise in fiber-optic signals [4].

In addition, Qu used the classical modal algorithm to decompose the vibration signal into multiple frequency band signals, and extracted the kurtosis of each frequency band as a feature for identification. However, the signal time domain of the same intrusion type may also have multiple vibration forms, which resulted in the number of frequency bands in which the algorithm decomposed the signal cannot be determined [5]. Sun proposed a method based on image morphology to extract temporal and spatial two-dimensional signal features, and the recognition rate is high, but the fiber pre-warning system needs to be in an ideal working environment. The environmental noise and interference have great influence on the recognition result [5].

To solve these problems, this paper innovatively introduces the combination of short-term analysis in time domain and wavelet analysis in wavelet domain. Based on the measured data, a multi-dimensional feature extraction and feature selection method which can fully characterize the vibration signal feature is proposed. Secondly, this paper innovatively preprocesses the signal using polynomial least squares method and wavelet threshold denoising method, which removes the trend term caused by the system, and reduces the influence of environmental noise. Then, based on QUICK-REDUCT algorithm to select features, remove redundant information, and extract the optimal feature vector combination under different dimensions. Finally, using the probabilistic neural network (PNN) for training and testing, and the recognition rate is as high as 92%.

2 Principle and Preprocessing

2.1 Eliminate Trend Terms Using Least Squares

The vibration signal data collected in the vibration test often deviates from the baseline, and even deviates from the baseline changing with time due to the zero drift of the amplifier with temperature changes, the instability of the low frequency performance outside the sensor frequency range, and the environmental interference around the sensor. The entire process of deviating from the baseline over time is referred to as the trend term of the signal. It can be seen from Fig. 1.

Set the time domain signal of a certain position of the fiber to $\{x_k\}$, $k = 1, 2, 3, \cdots, n$, where n is the number of sampling points. Since the sampling is equally spaced, the sampling interval can be set to 1. Use a polynomial function to represent the trend item of the signal:

$$\hat{x}_k = a_0 + a_1 k + a_2 k^2 + \ldots + a_m k^m = \sum_{j=0}^{m} a_j k_j, k = 1, 2, \ldots, n \tag{1}$$

Theoretical derivation shows that the trend term obtained when m = 0 is a linear trend term, eliminating the linear trend term and obtaining:

$$y_k = x_k - \hat{x}_k = x_k - (a_0 + a_1 k), k = 1, 2, \ldots, n \quad (2)$$

The curve trend term is obtained when $m \geq 2$. In actual speech signal processing, $m = 1 \sim 3$ is usually taken. In this paper, $m = 3$ is obtained to obtain the curve trend term, and then the trend term is eliminated by using polynomial least squares method for the fiber sensing signal.

Taking the artificial tapping signal of the field experiment as an example, the sampling frequency of the signal is 2 kHz, the signal length is 10 s, and the trend term is eliminated by using the least squares method. The signal waveforms before and after the trend item are shown in Figs. 1 and 2.

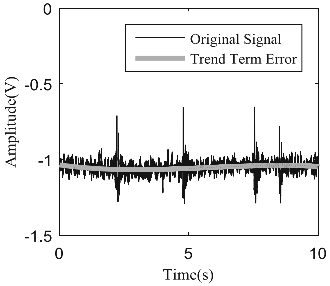

Fig. 1. Original signal curve

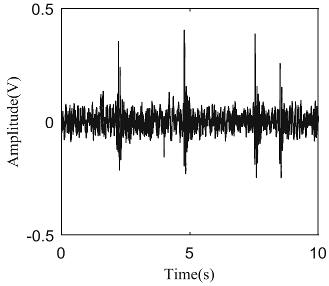

Fig. 2. Detrended term signal curve

2.2 Wavelet Threshold Method for Noise Reduction

Since the fiber sensing signal is susceptible to the background environment noise, how to retain the useful vibration signal to the utmost is a key problem in fiber optic vibration signal preprocessing. According to the characteristics that the vibration intrusion signal is a sudden signal, the conventional Fourier filter noise reduction method will simultaneously attenuate the vibration signal and the noise. However, wavelet analysis has a frequency zoom function, which can eliminate the sharp part of the signal and some weak small signals while eliminating noise. Therefore, the wavelet denoising algorithm is used to denoise the fiber vibration signal.

This section firstly uses a polynomial least squares method for a manual tapping signal with a length of 10 s to eliminate the trend error caused by the system to the fiber signal, Then, the wavelet threshold denoising is processed on the noisy fiber vibration intrusion signal after eliminating the trend term. The wavelet basis function used is db6 wavelet, and the number of decomposition layers is 7 layers. The waveform diagram before and after the vibration signal is denoised is shown in the figures below. Figure 3 is the waveform of the artificial tapping signal with no noise after eliminating the trend term. Figure 4 is the waveform of the artificial tap signal after noise reduction. Comparing the waveforms before and after noise reduction of the vibration signal, after the noise reduction by the wavelet threshold method, the high-frequency noise signal is obviously suppressed, and the sharp-changing part of the signal is better preserved, and the signal-to-noise ratio of the vibration intrusion signal is improved.

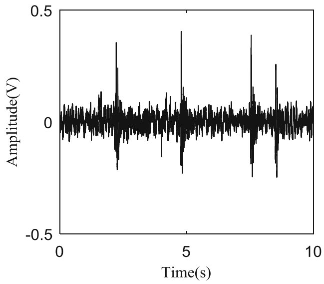

Fig. 3. Detrended term signal curve

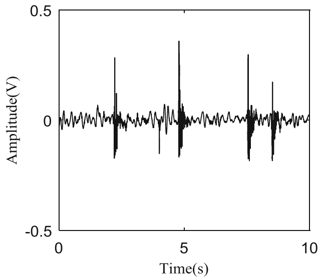

Fig. 4. Heuristic threshold denoising result

3 Feature Extraction and Identification

3.1 Time Domain Feature Extraction

After a large number of repeated tests, the fiber-sensing signals of different vibration intrusion types can be found that their time-domain characteristic parameters will satisfy certain statistical characteristics, and if the number of repeated tests tends to infinity, the feature distribution of the same type of signal is the same. Based on the differences in the distribution of statistical characteristics of different types of vibration signals, the time domain characteristic parameters of the signals can be used to identify the type of vibration intrusion.

Variance and Maximum Energy Segment. The variance reflects the fluctuation and dispersion of the one-dimensional signal. Vibration intrusion can cause fluctuations in the fiber's time domain signal, so the variance can be used to indicate the degree of fluctuation of the fiber's vibration signal. The greater the fluctuation of the time domain signal, the greater the variance of the signal; and vice versa. Variance can be expressed as

$$D(X) = Var(X) = E\{[X - EX]^2\} = \frac{1}{n}\sum_{i=1}^{n}(x_i - EX)^2 \tag{3}$$

Theoretically, the fluctuations of non-invasive signals, artificial tapping signals, vehicle passing signals, and mechanical excavation signals depend on the time of action of the vibration intrusion on the fiber. However, in practical applications, as is shown in Figs. 5, 6, 7 and 8, the fiber sensing signal is susceptible to the influence of the surrounding environment, so the signal collected by the system is mixed with a lot of high frequency noise.

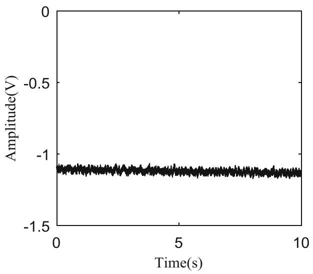

Fig. 5. No vibration intrusion signal

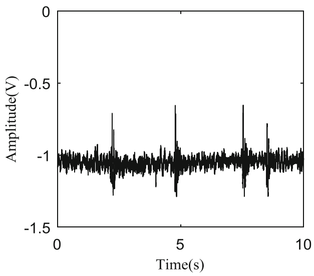

Fig. 6. Artificial tapping signal

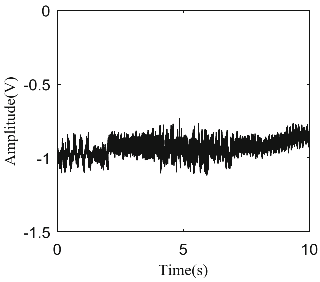

Fig. 7. Vehicle passing signal

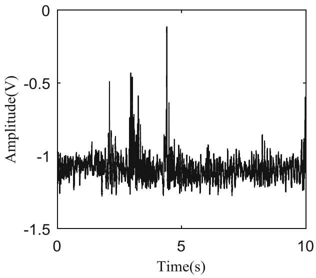

Fig. 8. Mechanical excavation signal

Based on a number of field experimental sample data with a time length of 10 s, the total energy of these four fiber signals is analyzed by observing the time domain waveforms of non-invasive signals, artificial tapping signals, vehicle passing signals and mechanical excavation signals. The duration of the vibration intrusion is different, and the duration of the vibration caused by the artificial tapping signal to the time domain signal is much less than 1 s; the vibration intrusion of the vehicle through the signal and the mechanical excavation signal takes a long time, which is reflected in the time domain signal. The vibration of several seconds continues to fluctuate, and there is a process of ascending, continuing, and descending. In order to minimize the influence of system noise, this section divides the sample data of 10 s into ten parts, each part is data with a time length of 1 s, and calculates the energy of ten time domain signals with a time length of 1 s, respectively. The larger of the ten energy values is the maximum energy segment characteristic of the time domain.

Short-term Average Amplitude. Fiber optic signals are unsteady, time-varying signals, but can be considered to be steady-state, time-invariant in "short-term" and "short-term" in milliseconds. According to this characteristic, the signal can be divided into several segments to analyze its characteristics, which is called short-term analysis technique of the signal, and each segment of the signal is called a frame signal.

The advantage of the short-term average amplitude is that it does not need to take the quadratic square, and the amplitude of the fiber signal is relatively small. Therefore, the short-term average amplitude is used to observe the signal. Firstly, the 10-s artificial tapping signal is processed by polynomial least squares method and wavelet threshold method, and then the short-term average amplitude method proposed in this section is used to analyze the short-term signal energy fluctuation. The frame processing is performed by Hanning window. The frame length is 30 ms, and the waveforms of the short-term average amplitudes of the four intrusion types are shown in Figs. 9, 10, 11 and 12.

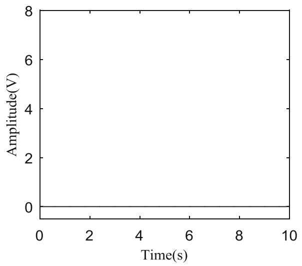

Fig. 9. Short-term average amplitude of vibration-free intrusion signal

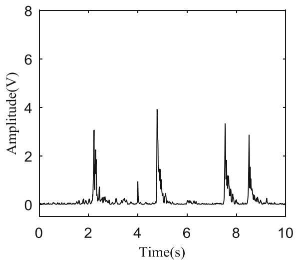

Fig. 10. Short-term average amplitude of artificial tapping signal

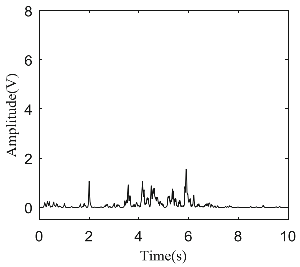

Fig. 11. Short-term average range of vehicles passing signal

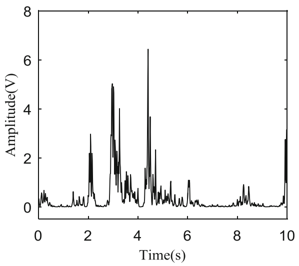

Fig. 12. Short-term average amplitude of mechanical excavation signal

It can be seen from the above figure that the high frequency noise of the original signal of the optical fiber is suppressed after calculating the short-term average amplitude, and the amplitude of the short-term average amplitude map of the non-invasive signal is almost constant to zero and does not change with time. In comparison, the short-term signal amplitude can express the amplitude characteristics of the fiber signal in the time domain.

The maximum segment of short-term average amplitude vibration is defined as the sum of the amplitudes of each vibration segment exceeding the threshold after the short-term amplitude analysis of the signal, and the maximum value is taken as the maximum segment of the vibration energy in all the vibration segments. The sum of short-term average amplitude vibrations is defined as the short-range amplitude analysis of the signal, and all the vibration segments exceeding the threshold are counted, and the sum of the amplitudes of all the vibration segments is obtained as the sum of the vibration energies.

3.2 Wavelet Domain Feature Extraction

In the foregoing, the characteristic parameters of the signal are analyzed from the perspective of time domain. The time domain features are mainly extracted by the amplitude variation. However, the characteristic parameters in a single domain cannot fully reflect the complex distribution of the intrusion signal. Therefore, this section from the wavelet domain The angle of the signal is analyzed. In this paper, discrete wavelet transform is used to analyze the signal in multiple resolutions, and the fraction of wavelet energy at each scale is extracted as the characteristic parameter of wavelet domain.

Considering that the Db wavelet has the characteristics of describing the non-stationary signal as a whole, combined with the results of multiple experiments [7], this paper selects the db6 wavelet and performs 7-layer wavelet decomposition on all the measured samples. The wavelet energy of each scale is then calculated and normalized to eliminate the effect of amplitude. After wavelet decomposition of the four types of samples, the statistical average results of the obtained wavelet energy ratio are shown in Figs. 13, 14, 15 and 16. The comparison shows that the energy ratios of different

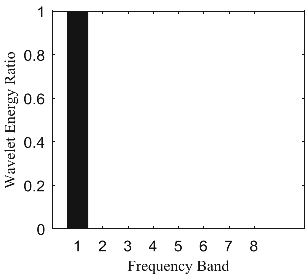

Fig. 13. Wavelet energy ratio of vibration-free intrusion signal

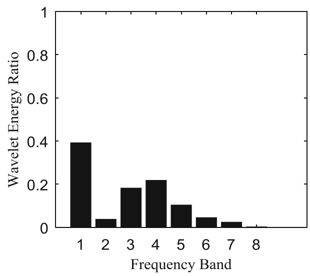

Fig. 14. Wavelet energy ratio of artificial tapping signal

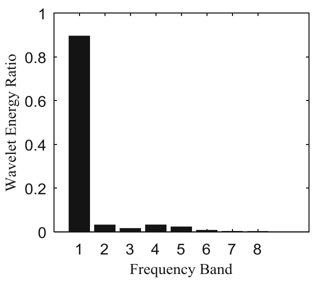

Fig. 15. Wavelet energy ratio of vehicle passing signal

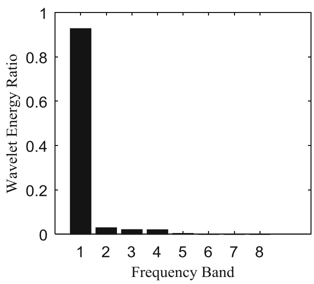

Fig. 16. Wavelet energy ratio of mechanical excavation signal

frequency bands without intrusion and manual tapping are significantly different. The distribution of vehicle passing and mechanical excavation is similar, but the difference between the characteristics of the first two intrusion signals is obvious.

4 Feature Evaluation and Identification

4.1 Feature Evaluation

In this paper, 12 eigenvalues are obtained after extracting the time domain and wavelet domain features of the four types of intrusion events, which are the maximum energy segment, the variance, the short-term average amplitude vibration maximum segment, the short-term average amplitude vibration sum, and the wavelet decomposition from the low frequency band to the high frequency band.

Jensen proposed the QUICKREDUCT algorithm for attribute reduction of rough sets [8]: the attribute set is initialized to an empty set, and attributes are added to it at a time. If this attribute is added, the attribute is retained and the attribute is retained. Otherwise, it is discarded. According to this idea, the design is as follows (Fig. 17):

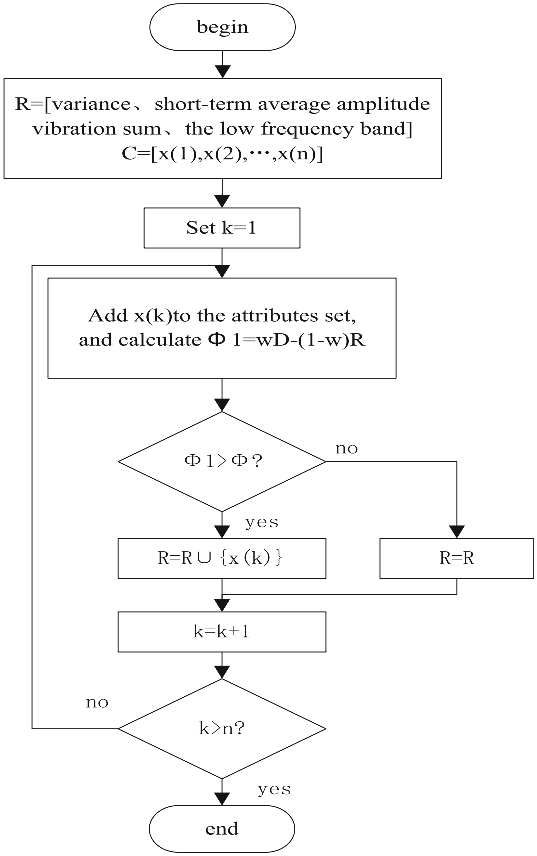

Fig. 17. QUICKREDUCT algorithm

Among them, x(k) is a variety of features, D is the correlation between the selected feature subset and the state category, and R is the redundancy between the various features.

Using the QUICKREDUCT algorithm, the optimal eigenvectors of different feature dimensions are shown in Table 1.

Table 1. Optimal eigenvectors of different dimensions

Dimension	Optimal feature vector combination
4	2 4 5 12
5	2 4 5 11 12
6	2 4 5 10 11 12
7	2 4 5 9 10 11 12
8	2 4 5 8 9 10 11 12
9	2 4 5 6 8 9 10 11 12
10	2 4 5 6 7 8 9 10 11 12
11	1 2 4 5 6 7 8 9 10 11 12
12	1 2 3 4 5 6 7 8 9 10 11 12

Through the above algorithm, the optimal feature vector corresponding to different feature vector dimensions is obtained. Compared with the exhaustive method, all the eigenvectors are calculated by the classifier. The method of this paper is to calculate the optimal eigenvectors corresponding to different dimensions, which can reduce the computational complexity.

4.2 Identification Based on PNN

Probabilistic neural network (PNN) consists of input layer, hidden layer, summation layer and output layer, and its structure is shown in the Fig. 18.

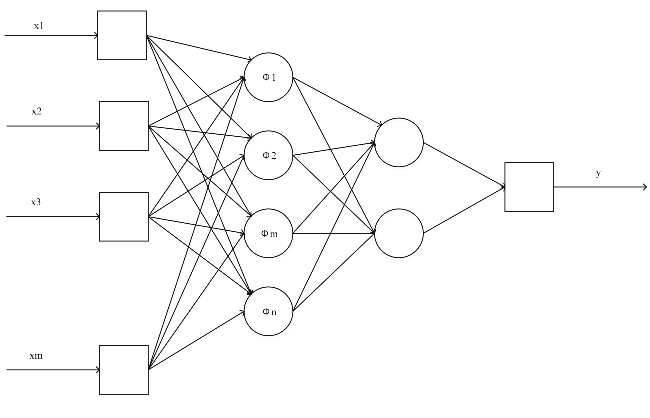

Fig. 18. Probabilistic neural network

Compared with the commonly used BP and radial basis neural networks, PNN has the advantages of simple network learning process, good expansion performance, few parameters and short training time. These advantages not only meet the real-time requirements of the system, but also reduce the impact of parameter settings on classification recognition [9]. Therefore, this paper chooses PNN neural network as the classifier of fiber-optic early warning system, and estimates the correct rate of classifier identification by double cross-validation method.

In general, the width parameter of the PNN, that is, the standard deviation is usually taken as 0.1. This paper uses 640 measured data samples to train tests. Among them, there are 196 groups of non-invasive sample data, 178 sets of manual tapping sample data, 126 sets of data after vehicle passing, and 140 sets of data for mechanical mining. Combined with the 5-fold cross-validation method, the average recognition accuracy of the optimal eigenvectors of different dimensions is obtained, as shown in Table 2.

Table 2. Average accuracy of identification of different eigenvalues

Feature vector dimension	Average recognition accuracy
4	92.8670%
5	93.7494%
6	92.8755%
7	93.7238%
8	92.9889%
9	93.2788%
10	92.8824%
11	92.6991%
12	93.7451%

5 Conclusion

Since fiber distributed vibration sensing (DVS) system based on phase-sensitive optical time domain reflectometer (Φ-OTDR) has the characteristics of identifying intrusion signals, wide monitoring range and high system sensitivity. The identification of vibration signal types is an important research topic. In this paper, the vibration signal is preprocessed, and the signal is preprocessed by using the polynomial least squares method and the wavelet threshold denoising method, which removes the trend term caused by the system, and reduces the influence of environmental noise. Secondly, this paper innovatively introduces the combination of short-term analysis in time domain and wavelet analysis in wavelet domain. Based on the measured data, a multi-dimensional feature extraction and feature selection method is proposed, which fully characterizes the characteristics of vibration signals. Finally, the feature selection is based on QUICKREDUCT algorithm, and the redundant information is removed to extract the optimal feature vector combination under different dimensions. Finally, using the probabilistic neural network (PNN) network for training and testing, and the recognition rate is over 92%.

Acknowledgement. Thanks to the support of Harbin Institute of Technology and Weihai Fund.

References

1. Ye, Q., Pan, Z., Wang, Z.: Progress of research and applications of phase-sensitive optical time domain reflectometry. Chin. J. Lasers **44**(06), 7–20 (2017)
2. Mahmoud, S., Visagathilagar, Y., Katsifolis, J.: Real-time distributed fiber optic sensor for security systems: performance, event classification and nuisance mitigation. Photonic Sens. **2** (3), 225–236 (2012)
3. Mahmoud, S., Katsifolis, J.: Robust event classification for a fiber optic perimeter intrusion detection system using level crossing features and artificial neural networks. In: SPIE Defense, Security, and Sensing. International Society for Optics and Photonics (2010)
4. Xu, C., Guan, J., Bao, M., et al.: Pattern recognition based on time-frequency analysis and convolutional neural networks for vibrational events in Φ-OTDR. Opt. Eng. **57**(1), 1 (2018)
5. Qu, Z., Li, J., Qi, S.: Signal analysis method for safe distributed optical fiber early warning system based on EMD. J. Tianjin Univ.: Nat. Sci. Eng. Technol. **40**(1), 73–77 (2007)
6. Sun, Q., Feng, W., Zeng, W.: Pattern recognition of optical fiber early warning system based on image processing. Opt. Precis. Eng. **23**(2), 334–341 (2015)
7. Sun, Q.: Research on pattern recognition method of Φ-OTDR optical fiber early warning system. Tianjin University (2015)
8. Jensen, R., Shen, Q.: Fuzzy-rough sets for descriptive dimensionality reduction. In: IEEE International Conference on Fuzzy Systems, vol. 1, pp. 29–34 (2002)
9. Hu, H., Hu, X., Guan, X.: Forecasting method of crude oil output based on optimization of LSSVM by particle swarm algorithm. In: International Conference on Information Science and Control Engineering. IEEE Computer Society, pp. 334–338 (2017)