Research on oil-gas Pipeline Leakage Detection Method Based on Particle Swarm Optimization Algorithm Optimized Support Vector Machine

Jianfeng Gao*,1, Yu Zheng 2, Kai Ni2, Huaizhi Zhang1, Bin Hao1, Jialin Yan2
1 School of Petrochemical Engineering & Environment, Zhejiang Ocean University, ZhouShan, ZheJiang, 316000, China
2 School of Ship & Maritime Transport, Zhejiang Ocean University, ZhouShan, ZheJiang, 316000, China
*Corresponding author’s e-mail: gaojf409@zjou.edu.cn

Abstract. In order to solve the problem of low accuracy in oil-gas pipeline leak detection, a pipeline leak detection method based on Particle Swarm Optimization (PSO) algorithm optimized Support Vector Machine (SVM) is introduced. This method uses PSO to solve the penalty factor ‘c’ and kernel function parameter ‘g’, and constructs the pipeline leakage detection model of SVM. We set up an experimental platform to collect negative pressure wave signals under different working conditions. After wavelet domain denoising and data preprocessing, four eigenvalues of Mean, Standard Deviation, Kurtosis and Skewness are extracted from the signals to form the eigenvector samples, which are taken as input of SVM, and four working conditions of normal, leakage, rise and fall are taken as output. Through experimental verification, the comprehensive performance of PSO-SVM algorithm is better than that of traditional SVM, Genetic Algorithm optimized SVM and grid search algorithm optimized SVM. The POS-SVM algorithm can be applied to the leak detection of oil-gas pipeline.

1. Introduction

By 2020, the scale of China's long-distance oil-gas pipeline network has reached 169,000 km. With the increase of the use time of oil-gas pipelines, the corrosion of pipeline, as well as the destruction of other natural and human factors, it is inevitable that oil-gas pipeline leakage accidents will occur. Oil-gas pipeline leakage accidents can cause environmental pollution, casualties and property losses. Therefore, how to efficiently detect oil-gas pipeline leakage has become a particularly important research topic[1].

It is called signal processing method that detecting the leakage status after collecting and analyzing the signals of oil-gas pipeline working by using various intelligent sensors. This detection method has the advantages of real-time detection and high applicability, However, due to the complexity of pipeline lines in reality, there are many noises, thus reducing the detection accuracy[2]. We introduce Support Vector Machine to train and learn the pipeline negative pressure wave signals collected after data preprocessing and intelligently identify them, and introduce Particle Swarm Optimization algorithm to optimize the recognition effect of SVM and improve the accuracy of pipeline leak detection[3].
2. Preliminaries

2.1. Support Vector Machine
Support Vector Machine is a machine learning method based on statistical learning theory invented by Vapnik et al in 1995[4]. The main idea of the algorithm is to map the sample set of the original space to a higher dimensional feature space through the kernel function, and then construct a hyperplane from this higher dimensional feature space, so that the original two types of data samples which are not linearly separable are as far away from the hyperplane as possible. The algorithm takes Structural Risk Minimization as the inductive principle, so it still has good generalization ability in the case of small sample and nonlinear learning. The model formula of support vector machine is shown in (1) and (2), The kernel functions commonly used in Support Vector Machine modeling are shown in Table 1.

\[
    f(x) = w^T \phi(x) + b \quad (1)
\]

\[
    \min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{n} \xi_i \quad (2)
    \text{ s.t. } y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i \quad i = 1, ..., n; \xi_i \geq 0
\]

Table 1. Formatting sections, subsections and subsubsections.

| Name               | Expression                                      |
|--------------------|-------------------------------------------------|
| Linear Kernel      | \(K(x_i, x_j) = x_i^T x_j\)                    |
| Polynomial Kernel  | \(K(x_i, x_j) = (\gamma x_i^T x_j + c)^d, \gamma > 0, d \leq 1\) |
| Gaussian Kernel    | \(K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\rho^2}\right), \rho > 0\) |

The performance of Support Vector Machine depends on the selection of kernel function, and there is no scientific and systematic method to solve the selection of kernel function. The most widely used kernel functions in scientific research are linear kernel and Gaussian kernel. In view of previous experience, if the characteristic dimension of the sample set is high and the number of samples is large enough, the selection of linear kernel can avoid a huge amount of computation; if the number of samples is not large enough and the characteristic dimension is small, the Gaussian kernel function is generally chosen[5]. The Gaussian kernel function has the advantages of wide mapping dimension, few parameters to be determined and relatively simple operation. Therefore, Gaussian kernel function is selected in the modeling of SVM in this paper.

The selection of penalty factor ‘c’ in Support Vector Machine modeling can achieve a balance between model complexity and training error, and the value depends on the specific problem; The selection of kernel function parameter ‘\(\gamma\)’ mainly reflects the range characteristics of training samples; ‘c’ and ‘\(\gamma\)’ directly affect the learning ability of the Support Vector Machine model[6].

2.2. Particle Swarm Optimization Algorithms
Particle Swarm Optimization algorithm is a population-based optimization algorithm proposed by American social psychologist James Kennedy and electrical engineer Russell Eberhart in 1995, inspired by the social behavior rules of birds and human beings[7]. This algorithm has the advantages of fast convergence, fewer parameters to be set, simple implementation and so on, and has been widely used in many fields in recent years[8].

The algorithm can be described as: it is assumed that in an n-dimensional space, there are m particles and each particle is a feasible solution of the problem to be optimized in the search space[9].
Each particle can remember its own search for the optimal solution, and the optimal position experienced by the entire particle swarm, that is, the global optimal solution found so far. Moreover, each particle has an initial velocity, and when both optimal solutions are found, they both update their velocities according to Equation (3).

$$\begin{align*}
V_{id}(t+1) &= \omega V_{id}(t) + c_1 r_1 (p_{id} - z_{id}(t)) + c_2 r_2 (p_{gd} - z_{id}(t)) \\
z_{id}(t+1) &= z_{id}(t) + v_{id}(t+1)
\end{align*}$$

(3)

Among them:
- \(z = \{z_1, z_2, z_3, \ldots, z_m\}\): Population named ‘z’ consisting of m particles;
- \(p_{id}\): Individual optimum;
- \(p_{gd}\): Global optimum;
- \(c_1, c_2\): Acceleration factor;
- \(\omega\): Inertia coefficient;
- \(r_1, r_2\): A random number between 0 and 1.

Through the above, we can summarize the program frame diagram of Particle Swarm Optimization algorithm optimized Support Vector Machine algorithm, as shown in Figure 1:

Figure 1. Program frame diagram of Particle Swarm Optimization algorithm optimized Support Vector Machine algorithm

3. Experimental analysis

3.1. Data acquisition

The author built an experimental platform for pipeline leakage to study the negative pressure wave signals under different working conditions during pipeline operation. The experimental platform is shown in Figure 2:

The experimental system mainly consisted of air compressor, gas buffer tank, separation tank, pressure reducing valve, pressure gauge, flow meter, various specifications of ball valve and different specifications of steel pipe[10].
Figure 2. The experimental platform

Considering the safety of the experiment, compressed air was used as the experimental medium instead of natural gas in this study. Considering the convenience of the experiment, a ball valve with an opening of 0.2cm was used to simulate the pipeline leakage, Chengdu TESTER CY301 high precision intelligent pressure sensor was used to collect negative pressure wave signal, this sensor can be directly connected with the personal computer to monitor and record pressure changes in real time and display the curve of pressure over time. The sensor picture is shown in Figure 3.

Figure 3. CY301 high precision intelligent pressure sensor

Three leakage points were selected for the experiment. The sampling frequency of the sensor was set as 1000Hz, and the changes of negative pressure wave signals under the pipeline conveying pressure of 0.5MPa, 0.8MPa and 1MPa were collected under 7 working conditions. Among them, the 7 working conditions are: normal working condition, valve is being turned up, leakage working condition, impact working condition, compressor is being closed, valve is being turned down, compressor is being opened. Since the change trend of negative pressure wave shape was basically the same under the same working condition but different conveying pressure, we only shown the change of negative pressure wave under different working conditions when the conveying pressure is 0.5MPa. The waveform diagram is shown in Figure 4:

From the above negative pressure wave changes, we can conclude that: under normal working conditions, the pipeline negative pressure waveform will tend to be stable; When the valve is being turned up, or the compressor is being opened, or under the impact condition, the negative pressure waveform will rise; When the valve is being turned down, or the compressor is being closed, the negative pressure waveform will decline. When the pipeline is leaking, the negative pressure waveform will show a small decline at the moment of leakage, and then the waveform will tend to be stable.
(a) Under normal working conditions

(b) When the valve is being turned up

(c) Under leakage working condition
(d) Under impact working condition

(e) When the compressor is being closed

(f) When the valve is being turned down
3.2. Data preprocessing

Because the pipeline leakage is sudden, and the pressure in the pipeline often cannot change suddenly, there are many noises in the experimental data, which will affect the capture of the sudden descent point, and then affect the final detection accuracy. So, we used method of wavelet domain denoising to denoise the negative pressure wave signal.

Using the wavelet analysis toolbox in MATLAB, we selected the 'db5' wavelet basis function to decompose the signal in five layers, and used the ‘Minimax’ threshold to denoise the negative pressure wave signals collected in the experiment under leakage conditions. The experimental results are shown in Figure 5:

In order to simplify the training complexity of the Support Vector Machine model and predict the actual leakage conditions more accurately, according to the above negative pressure wave shape changes, the above 7 conditions were integrated into 4 conditions, namely: 1-normal condition, 2-leakage condition, 3-rising condition and 4-decline condition. Thirty-four experiments were repeated for each group of working conditions, and a total of 136 groups of negative pressure wave change data were obtained. The Mean, Standard Deviation, Kurtosis and Skewness of the 136 groups of negative
pressure wave signals were respectively calculated as the input feature vectors of Support Vector Machine, and four working conditions of 1-normal, 2-leakage, 3-rise and 4-fall are taken as output[11].

We calculated that the Mean of negative pressure wave signals under four working conditions was mainly distributed between 4 and 13, Standard Deviation was distributed between 0 and 1, Kurtosis was distributed between 0 and 8, and Skewness was distributed between -2 and 3. In order to eliminate the difference between different dimensions of the eigenvalues, avoid a single variable playing a leading role in the influence of the target variable, and simplify the inner product operation between the eigenvectors of the kernel function; We conducted data normalization processing on the data set to make each coefficient in the data set scale linearly to between [-1, 1]. The data normalization formula (4) used in this paper is shown as follows:

\[
Y = -1 + \frac{1 - (-1)}{x_{max} - x_{min}}(x - x_{min})
\] (4)

3.3. Identification and Analysis
Libsvm toolbox in MATLAB was used to predict and identify different pipeline working conditions. The key parameters ‘c’ and ‘g’ of Support Vector Machine were optimized by three different optimization methods, namely: Genetic Algorithm, Particle Swarm Optimization algorithm and grid search algorithm. The test accuracy and operation time of different optimization algorithms were compared, which were taken as the evaluation index of the algorithm’s merits[12]. The program running results of the three optimization algorithms are shown in Table 2:

In the table, we abbreviate the Particle Swarm Optimization algorithm as PSO, the grid search optimization algorithm as GSO, and the Genetic Algorithm as GA.

| Algorithm | Number of test sets | Accuracy(%) | Running time(second) |
|-----------|---------------------|-------------|----------------------|
| PSO       | 18                  | 94.4444     | 0.99                 |
| GA        | 18                  | 88.8889     | 3.98                 |
| GSO       | 18                  | 94.4444     | 9.49                 |
| PSO       | 28                  | 96.4286     | 1.01                 |
| GA        | 28                  | 92.8571     | 4.07                 |
| GSO       | 28                  | 100         | 9.91                 |
| PSO       | 38                  | 92.1053     | 1.10                 |
| GA        | 38                  | 89.4737     | 3.52                 |
| GSO       | 38                  | 92.1053     | 8.96                 |
| PSO       | 48                  | 91.6667     | 0.98                 |
| GA        | 48                  | 83.3333     | 2.99                 |
| GSO       | 48                  | 93.75       | 8.51                 |

It was concluded that, in terms of accuracy: grid search algorithm > Particle Swarm Optimization algorithm > Genetic Algorithm; In terms of the length of operation time: Particle Swarm Optimization algorithm < Genetic Algorithm < grid search algorithm.

The core idea of grid search algorithm is exhaustive attack method. This algorithm avoids the precocity phenomenon of the heuristic algorithm, and may be better than the heuristic algorithm in results, such as the Genetic Algorithm and Particle Swarm Optimization algorithm in this paper. However, when the number of training sets or the dimension of feature vectors or the parameters to be optimized increase, the time complexity of the algorithm will increase exponentially, which greatly increases the running burden of the program.
We changed the total number of samples to 280, and then compared and analyzed the three algorithms. The results are shown in Table 3:

Table 3. The prediction accuracy and running time of different algorithms when the data set is 280

| Algorithm | Number of test sets | Accuracy(%) | Running time(second) |
|-----------|---------------------|-------------|----------------------|
| PSO       | 56                  | 94.6429     | 0.98                 |
| GA        | 56                  | 91.0714     | 20.83                |
| GSO       | 56                  | 96.4286     | 33.82                |

In the actual long-distance oil-gas transmission pipeline, the data of pipeline operation condition is far more complex than the data obtained in this experiment, and the data samples are far more than the samples obtained in this experiment. The dimension of the characteristic vector of the negative pressure wave signal is also more than the dimension of the characteristic vector extracted in this experiment. If the prediction accuracy is deliberately pursued, it will bring a great burden to the program operation and hardware equipment, which is not conducive to the maintenance of equipment and the timely discovery of leakage points. It can be seen from the experimental analysis that the average prediction accuracy of the Particle Swarm Optimization algorithm is only 1.488% lower than that of the grid search algorithm. The average prediction accuracy of the three algorithms is shown in Table 4. Therefore, considering the prediction accuracy and running time comprehensively, the pipeline leak detection method based on Particle Swarm Optimization algorithm can be applied to the leak detection of oil-gas pipelines.

Table 4. The average prediction accuracy of the three algorithms

| Algorithm | Accuracy(%) |
|-----------|-------------|
| PSO       | 93.85758    |
| GA        | 89.12488    |
| GSO       | 95.34566    |

4. conclusion
The innovation points of this study are as follows:

(1)An experimental platform was built to study the leak detection of oil-gas pipelines, which explored more possibilities for the study of the leak detection of oil-gas pipelines;

(2)The advantages and disadvantages of the three different algorithms were comprehensively compared from the two aspects of recognition accuracy and program running time. Finally, it was concluded that the pipeline leak detection method method Based on Particle Swarm Optimization Algorithm Optimized Support Vector Machine is more suitable for the actual oil and gas pipeline leak detection.

The shortcoming of this study is that: There are some differences between the experimental data and the actual pipeline operation data, and it still needs to be further verified whether this method can completely solve the problem of false positives and missed positives in the actual pipeline[13].

Acknowledgements
This research was funded by fund project. Fund project: Zhoushan Science and Technology Bureau research project "Research on the Leak Detection and Monitoring System of Submarine Oil and gas Transmission Pipeline Based on Wavelet Analysis", 2017C41004.

References
[1] Kan Z, Lang X.M., Wang X.L. (2018) Leakage Location of Branch Pipeline Based on Cyber-physical System Architecture. Information And Control., 47.01: 22–28.
[2] Song Y.J., Li S.Z. (2021) Gas leak detection in galvanised steel pipe with internal flow noise using convolutional neural network. Process Safety and Environmental Protection., Volume 146: 736-744.

[3] Xiao R., Hu Q.F, Li J. (2019) Leak detection of gas pipelines using acoustic signals based on wavelet transform and Support Vector Machine. Measurement., Volume 146: 479-489.

[4] Li M.H. (2017) Research on leakage detection and leakage point location of long distance oil pipeline. Dalian University of Technology., MA Thesis.

[5] Du C., Shao J.H., Yang W., Wang Z.S., Deng L.J., Shen H.J. (2021) Support vector machine indoor visible light positioning optimized by grid search method. Chinese Journal of Laser., 42(03): 104-109.

[6] Liu J.H, Su H.G., Ma Y.G., Wang G., Wang Y., Zhang K. (2016) Chaos characteristics and least squares support vector machines based online pipeline small leakages detection. Chaos, Solitons and Fractals., Volume 91: 656-669.

[7] Yang L., Ni L. (2015) Leak Location of Pipelines Based on IPSO-SVM. Journal of Changshu Institute of Technology., 29(02):76-81.

[8] Wang Y. (2020) Research on Oil Depot Information System Technology Based on Quantum Particle Swarms Optimization. Zhejiang Ocean University., MA Thesis.

[9] Wang C., Han F., Zhang Y., Lu J.Y. (2020) An SAE-based resampling SVM ensemble learning paradigm for pipeline leakage detection. Neurocomputing., Volume 403: 237-246.

[10] Zhou S.T. (2019) Research on Submarine Oil and Gas Pipeline Leak Detection and Monitoring System Based on Wavelet Analysis. Zhejiang Ocean University., MA Thesis.

[11] Ji J., Hu H., Gao Y., Shao X.c., Xuan J. (2020) The Wind Power Prediction Based on The Genetic Algorithm to Optimize Parameters of Support Vector Machine. Electronic Measurement and Testing., 2020(21): 32-35.

[12] Yan Z.W., Li P., Lang X.M., Liu H.Y. (2020) Research on Pipeline Leakage Detection Method Based on Genetic Algorithm Optimized BP Neural Network., Contemporary Chemical Industry., 49,01: 216-220+227.

[13] Gao J.F., Wang Y., Jin J.H. (2020) Fire warning algorithm based on QPSO-BP neural network. Fire Science and Technology., 39(10): 1345-1349.