Research on Pipeline Defect Classification Based on SVM

Peisheng Liu, Yunping Han* and Chengshun Dong

The Computer and Communication Engineering College, Liaoning Petrochemical University, Fushun Liaoning, 113001

Corresponding author’s e-mail: *hanyunping@lnpu.edu.cn

Abstract. The defect identification and evaluation of buried steel pipeline is a long-term challenge in the field of pipeline detection, and the prerequisite for efficient identification of defects is the accurate extraction of pipeline damage signals. Aiming at the characteristics of buried pipeline defect signals, a method of pipeline defect signal extraction and recognition is proposed, which is based on support vector machine (SVM). A dictionary is obtained by learning from the original signal, the dictionary is used to construct a sparse model of the defect signal using a regularized orthogonal matching pursuit algorithm, and the feature vector of the signal is obtained according to the compressed sensing theory. Furthermore, multi-classification SVM is used to establish a mapping relationship between the feature vector of the defect signal and the actual defect type of the pipeline, and Genetic Algorithm-Particle Swarm Optimization is used to guide the selection of SVM parameters. The results showed that the proposed classification method can realize the accurate division of the damage degree of pipeline defects.

1. Introduction

Buried pipeline is one of the most effective means of oil and gas transportation. Due to the influence of corrosion, third-party damage and natural disasters, some damages will inevitably be formed. It is necessary to timely detect defects through relevant detection methods and assess their impact on pipeline safety [1]. Conventional non-destructive testing technology is commonly used in pipeline detection with ultrasonic testing and magnetic flux leakage testing. The traditional internal inspection technology applies only to have formed the macroscopic defects [2]. In addition, most of the buried pipeline has the characteristics of limit pigging, therefore in the condition of trenchless, pipeline defect detection is a problem to be solved.

At present, some of the external detection techniques available include eddy current method, the guided wave detection, transient electromagnetic method [3] and the method of X-ray scan, the above method is called the active detection. But the above methods are external electromagnetic excitation method, increasing the difficulty of field detection, and the precision for pipeline damage grades of a larger problem.

Relevant scholars conducted experiments and studies based on the relationship between defect magnetic signals and defect parameters, aiming at the limitations of metal magnetic memory detection criteria for pipeline defect. However, the pipeline defect identification methods used in relevant studies have some problems, such as fewer experimental samples, insufficient universality of the identification model, and lack of field actual verification. In this paper, based on sparse model and pipeline defect damage degree of the SVM classification model, by using the sparse model to extract the essence of pipeline defect feature vector, and the defect feature vector by improving the classification support...
vector machine (SVM) classification, for buried steel pipeline defect detection in the condition of trenchless pipe body injury provides an effective method.

2. Sparse model for feature extraction of pipeline damage signal

2.1. Eigen modal function basis dictionary learning

The pipe defect signals obtained through experiments can be decomposed into intrinsic mode function (IMF) using VMD algorithm. IMF is a composite signal of sinusoidal components with different frequencies. The dictionary \( D=[d_1; d_2; \ldots; d_N] \in \mathbb{R}^{N \times M} \) element can be designed to be a signal with different frequencies, different phases and different delays.

2.1.1. Dictionary initialization method. In order to ensure the rapid convergence of dictionary learning, make different atoms have a small degree of correlation, and improve the sparse modeling ability of dictionary learning, K-means algorithm is adopted to classify the training set composed of eigenmode functions.

Firstly, the correlation between the intrinsic modal functions contained in the training sample is calculated. The correlation can be expressed as:

\[
\mu = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} \langle y_i, y_j \rangle}{(I-1)(J-1)} \quad (i \neq j)
\]

(1)

The relevancy of learning dictionary refers to the maximum value of the relevancy between each atom in the dictionary, which can be expressed as:

\[
\mu = \mathcal{D} = \max_{i \neq j} \left\{ \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} \langle y_i, y_j \rangle}{(I-1)(J-1)} \right\}
\]

(2)

Secondly, calculate the average correlation between each eigenmode function and the other eigenmode functions \( \mu \).

Thirdly, assuming that the dimension of the dictionary is \( D \in \mathbb{R}^{N \times M} \), the M eigenmode functions with the lowest average correlation were regarded as the initial clustering center. If in this M intrinsic mode function, the correlation of any two eigen modal functions is higher than the mean of their respective correlations, then the eigen modal function with the least correlation of the other eigen modal functions replaces the eigen modal functions of the above two eigen modal functions, and the above process is repeated until convergence.

Fourthly, the atom obtained above is taken as the initial clustering center, and further clustering is carried out with \( \mu \) as the measurement standard, and K-means is adopted as the clustering algorithm, and the final clustering center is taken as the atom in the initial dictionary \( D \).

Finally, the dictionary is initialized to update all the dictionaries.

2.1.2. Dictionary update method for K singular value decomposition. In order to prevent the inverse operation of the matrix when the dictionary is updated, the K-singular value decomposition algorithm is adopted in this paper. When K singular value decomposition algorithm updates the dictionary according to columns, its corresponding sparse coefficient synchronization is finer, which is more conducive to algorithm convergence. Update each column atom \( d_j \) and its corresponding coefficient \( x_{j0} \), where, \( x_{j0} \) represents the \( j \)th line of matrix X, and the solution formula can be expressed as:

\[
\|Y - AD\|_2^2 = \left\| Y - \sum_{j=1}^{S} a_j d_j^T \right\|_2^2 = \left\| Y - \sum_{j=1}^{S} a_j d_j^T - a_k d_k^T \right\|_2^2
\]

(3)

Where, \( Y \) is the training set, \( A \) is the set of sparse models, \( d_j \) is the \( j \)th atom in dictionary \( D \), \( a_k \) is the \( k \)th line in model set \( A \).
Let $E_{ji} = \left(Y - \sum_{j=0}^{n-1} a_{ji} d_j\right)$ represent the error generated by the data in the training sample when the atoms are not reconstructed, so that the optimal solutions of formula (3) are minimized as $a_{j0}$ and $d_{j0}$. Finally, the error caused by different units of measurement can be avoided by normalizing the atoms in the dictionary.

2.2. Sparse model

In this paper, an iterative threshold contraction algorithm based on greedy algorithm theory is presented, which can accurately represent pipeline defect signals. In order to efficiently and rapidly process the information in the dictionary, the normalized orthogonal matching tracking algorithm is adopted to divide the correlation coefficients of atoms in $D$ into several groups:

$$j = \arg \min_{i=1,2,...,M} \left| g_i \right|$$

(4)

In the above group, the inner product of each column vector in the group and its residual is calculated, and the data whose absolute value is less than two times of the minimum value is selected and put into the support set to ensure that the support set has the characteristic of energy maximization.

The pipeline defect signals are represented as IMF linear combination obtained by VMD decomposition, and the modal signals obtained were approximately orthogonal, so the sparse model could not be obtained directly through the decomposed signals. In hierarchical sparse model, each component signal has its corresponding dictionary, so the component signals can be sparsely modelled separately, and then the dictionary of mixed signals can be formed by the sub-dictionary of component signals, so the signals can be sparsely processed effectively and accurately.

In hierarchical sparse model, regular terms $\lambda_2 \sum_{i=1}^{g} \left\| v_{[g_i]} \right\|_2$ need to be introduced, so that the sparse model is sparse in both dictionaries $D$ and its sub-dictionaries $D_i$:

$$v^* = \arg \min_{v \in \mathbb{R}^n} \left( \frac{1}{2} \left\| Y - Dv \right\|_2^2 + \lambda_2 \sum_{i=1}^{g} \left\| v_{[g_i]} \right\|_2 + \lambda_1 \left\| v \right\|_1 \right)$$

(5)

Where, $v_i$ is the child vectors of vector $v$, representing the corresponding coefficients of atoms in dictionary $D$ and dictionary $D_i$.

In the sparse model of pipeline defect signal, the sparsity of pipeline defect signal needs to be constrained to obtain a model with lower complexity than that based on the hierarchical sparse model theory. In the proposed method of pipeline defect signal model based on eigen modal function basis dictionary, pipeline signal model based on regularized orthogonal matching tracing can ensure the accuracy and computational efficiency of sparse model.

According to the actual pipe signal characteristics, a simulation signal is constructed as follows:

$$y = \cos(10x + 1) + e^{\sin(3x)}$$

(6)

Orthogonal matching tracing in sparse model is based on greedy algorithm theory, the optimization problem is transformed into searching $\left\| Y - Dv \right\|_2^2$ and the maximum atomic inner product problem in dictionary $D$, and the global optimal solution is obtained through iteration. Firstly, choose an atom to match signal $y$, build a sparse approximation, and find out the residual. Secondly, select the atoms most relevant to the residuals and put them into the support set. Then Schmidt transformation is performed on the atoms in the set, and then signal $y$ is mapped to the updated dictionary. The component of signal $y$ in the set is obtained, and the component is subtracted from the signal $y$ to obtain the residual signal. Repeat the above process and continue to decompose the residual signal until the pre-set condition is reached.
2.3. Signal feature extraction method based on compressed sensing theory

After signal sparse model, signal sampling and compression synchronization are realized through compressed sensing method processing. For pipeline defect signal compression perception, it can be divided into sparse model, compression processing and signal reconstruction. After the sparse signal is mapped into the corresponding matrix through compression processing, the resulting matrix is the measurement matrix, and the measured value can be expressed as:

$$y = \Omega u$$  \hspace{1cm} (7)

Where, $\Omega \in \mathbb{R}^{T \times N}$ is measurement matrix, $T$ is measuring frequency, $N$ is the length of the signal.

Based on IMFBD, the measurement signal sparse model is presented as follows:

$$x = Du$$  \hspace{1cm} (8)

By formula (7) and (8), the measured value can be expressed as:

$$y = \Omega u = \Omega D^T x = \tilde{\Omega} x$$  \hspace{1cm} (9)

Where, $\tilde{\Omega} \in \mathbb{R}^{T \times N}$ is the perception matrix.

The extraction process of magnetic anomaly signal feature in sparse expression of eigen modal function basis dictionary is as follows. Firstly, a dictionary study of the IMF is carried out. Secondly, a sparse model of magnetic anomaly signal based on IMFBD is established, and a dictionary learning method of IMFBD is given, including dictionary initialization, sparse model based on iterative threshold contraction algorithm and dictionary updating based on K singular value decomposition. Finally, on this basis, a method of magnetic anomaly signal sparse model based on the idea of hierarchical sparse model is proposed by using the idea of hierarchical sparse model.

The sparse model is determined by its basis function. After the basis function is determined, the perception matrix is measured, and then its measured value is obtained. In the process of signal reconstruction, each element of measured value is equally important. Therefore, missing one or several elements can still reconstruct pipeline defect signals and solve the problem of classification and identification of characteristic signals.

3. Classification of tubular damage based on support vector machine

3.1. Construction of multi-classification support vector machine

The selected kernel functions of the multi-classification support vector machine constructed in this paper are composed of polynomial kernel function and RBF kernel function. In order to enable support vector machine to have local and global optimal performance, the weight factor ($\eta$) is introduced to enable it to reasonably select kernel function in different situations and determine the reasonable weight factor through the input training sample data so as to construct a multi-classification SVM. The expression form is as follows:

$$K_{mix}(X,Y) = \eta K_g(X,Y) + (1-\eta) K_l(X,Y)$$  \hspace{1cm} (10)

Where, $K_{mix}(X,Y)$ is kernel function of multi-classification SVM, $K_g(X,Y)$ is the global kernel function, $K_l(X,Y)$ is the local kernel function, $\eta$ is weighting factor with $0<\eta<1$.

3.2. Pipeline defect classification model based on GAPSO-SVM algorithm

The parameter selection of SVM has an important influence on its classification accuracy and learning speed [4]. Unreasonable parameters will lead to SVM falling into local optimum or failing to get expected results. Due to the shortcomings of SVM algorithm, this paper combines genetic algorithm and particle swarm optimization algorithm to build a GAPSO optimization model, and takes their respective advantages to form a new hybrid intelligent algorithm [5].

The termination principles of GAPSO algorithm are as follows: (1) Achieve the optimized iteration times; (2) The tolerance value of the fitness function is less than its pre-set value for 60 consecutive times, that is, the global optimal solution. The selection of inertia weights in the initial parameters set
parameters is extremely critical. Large inertia parameters have good global search capability, and the low inertia weight corresponds to the local searching ability. Therefore, with the increase of optimization iterations, \( \kappa \) gradually decreases to ensure that the algorithm has good capability of global optimization and a good ability of local optimization in the later. In this paper, \( \kappa \) decreases linearly from 0.9 to 0.3 with the number of iterations, that is, \( \kappa_i = 0.9 - 0.6i/\text{max.num} \). In addition, particle swarm size is set as 20, evolutionary algebra is set as 200, crossover probability is set as 0.3, and mutation probability is set as 0.015. In order to balance the influence of random factors, the learning factor is set as \( c_1 = c_2 = 1.5 \).

In the GAPSO-SVM model, GAPSO algorithm is used to optimize the penalty parameter \( c \), weight factor \( \eta \) and kernel function \( \sigma_1 \) and \( \sigma_2 \), and the initial particle swarm vector \( Y \) is constructed:

\[
Y = [\eta, c, \sigma_1, \sigma_2]
\]  

(11)

In order to obtain good classification performance of SVM, the fitness function \( F(y_i) \) of the classification accuracy \( E(y_i) \) of its training data set is

\[
F(x_i) = E(x_i)
\]  

(12)

Where, \( F(x_i) \) is the fitness value of the \( i \)th individual particle and \( E(y_i) \) is the classification accuracy of SVM corresponding to the \( i \)th individual particle.

In order to ensure that the population searches within a reasonable range, the location and velocity are fixed within a pre-set interval, that is, the values of \( c \) and \( g \) are [0.1, 100].

### 4. Analysis of experimental data

The experimental tube is made of Q235 steel with a wall thickness of 3 mm and a diameter of 75 mm. In order to study the magnetic abnormal signals of different defects, 6 defects are made at different positions of the pipeline, namely, axial groove, transverse groove, rectangular groove, through hole, blind hole and 45° groove. The lifting distance during the experimental detection is 5 times of the diameter of the pipeline.

In order to obtain training samples of SVM and verify sample data sets, defect identification tests are carried out on experimental pipelines, and the damage degree of pipeline defects was divided into three types: low risk, medium risk and high risk. The data is shown in Figure 1.

![Figure 1. Pipe defect characteristic information table.](image)

After the sparse model of the detection data, the sample data and the pipe defect characteristics are marked, and the feature vector of the pipe body defect signal is used to represent different damage degrees. 100 groups of samples are collected for each damage level with a total of 300 groups of samples. The first 50 groups of each level are taken as training samples, and the last 50 groups are taken as test samples. According to the correspondence between the defect signals of training samples and the damage degree of pipeline defects, the normalized gradient energy index is divided into three sections, namely [0,0.2], [0.2,0.6] and [0.6,1.0], corresponding to the actual damage degree of pipeline defects as...
low risk, medium risk and high risk, respectively. Table 1 shows the comparison of classification accuracy under different parameters.

| Group | η   | c   | σ₁  | σ₂  | Classification accuracy /% |
|-------|-----|-----|-----|-----|----------------------------|
| 1     | 0.1 | 1.0 | 0.5 | 0.1 | 76.60                      |
| 2     | 0.3 | 3.0 | 1.0 | 0.3 | 72.49                      |
| 3     | 0.5 | 30.0| 2.0 | 0.5 | 92.36                      |
| 4     | 0.7 | 50.0| 5.0 | 1.0 | 86.28                      |
| 5     | 0.9 | 100.0| 10.0| 2.0 | 82.61                      |

In the process of optimizing SVM parameter selection based on GAPSO algorithm, it is assumed that the weight factor η is [0,1], the kernel function parameters σ₁ and σ₂ take the range of [0,10], the penalty parameter c is [0,100], the population number is determined to be 20, and the maximum number of iterations is 200. In order to avoid the occurrence of random error, GAPSO algorithm is used to optimize the SVM parameters are calculated respectively 10 times, penalty parameter value is determined by GAPSO 30.28, weighting factor η for 0.4 and the kernel parameter σ₁ is 2.63 and σ₂ is 0.96. The optimized parameters are substituted into the SVM composed of polynomial kernel function and Gaussian radial basis kernel function to construct the pipeline defect damage degree diagnosis model, and then the test samples are analysed.

The results show that the accuracy of the GAPSO-SVM model is 93.93%, higher than that of the SVM model without parameter optimization, and the calculation results are more stable each time. Therefore, the GAPSO-SVM model is applied to the identification and classification of buried pipeline defects.

5. Conclusions

In this paper, a classification model based on sparse model and SVM is proposed for the classification of trenchless defects in buried pipelines, and the defect recognition capability of the proposed classification model and other models is analysed. This paper draws the following conclusions:

1) In order to accurately extract the essential characteristics of pipeline damage signals, this paper proposes a sparse model method for pipeline defect signals based on the eigen modal function basis dictionary. On the basis of the above work, the sparse model of pipeline defect signal is analysed through compressed sensing algorithm to extract its characteristic coefficient, and the signal reconstruction, namely feature extraction, is verified and analysed through experiments.

After the feature extraction of the pipeline defect signal sparse model is completed, the multi-classification support vector machine based on GAPSO is used to classify the pipeline defect damage degree. According to the characteristics of GAPSO algorithm rapid global optimization and support vector machine structure risk minimization, the pipeline defect damage degree evaluation model based on sparse model and SVM algorithm is constructed. The experimental analysis shows that the classification method proposed in this paper has a high accuracy and certain advantages in the stability and accuracy of decision-making, and can effectively solve the problem of classification of buried pipeline damage.

References

[1] Li Z, Jarvis R, Nagy P B, Dixon S and Cawley P 2017 NDT & E Int. 92 59.
[2] Liu B, He L Y, Zhang H, Yang C and Henrique F 2017 Measurement 103 275.
[3] Hu B, Yu R and Liu J 2016 Anti-Corros. Method. M. 63 262.
[4] Lei N, Jiang J and Yong P 2013 J. Loss Prevent. Proc. 26 1085.
[5] Duisterwinkel E H A, Talnishnikh E, Krijnders D and Wortche H J 2018 IEEE T. Instrum. Meas. 67 655.