

Neural Network and Data Fusion in the Application Research of Natural Gas Pipeline Leakage Detection

Bingkun Gao¹, Guojun Shi^{1,2} and Qing Wang¹

¹ School of Electrical Engineering & Information
Northeast Petroleum University, DaQing, China

² School of Information Technology
Heilongjiang BaYi Agricultural University, DaQing, China

¹{gaobk, wqq8392} @163.com

Abstract

For natural gas pipeline, it has a leak or not is critical. The most commonly problems in the pipeline leak detection methods are the difficulties to identify, inaccuracy to locate, thus, the natural gas pipeline detection is difficult to be applied, therefore, the use of neural network multi-sensor data fusion of the natural gas pipeline leak detection is particularly important. In this paper, the method is proposed based on RBF neural network and the data fusion of D-S evidence theory for detecting the pipeline leak. Extracting neural network's input parameters through wavelet denoising, then substitute them into neural network and calculate them by multi-sensor data fusion algorithm so as to acquire leaking information.

Keywords: Leak detection; RBF neural network; Wavelet denoising; Data fusion; D-S evidence theory

1. Introduction

With the expansion of network gathering system, it is a tendency for gas pipeline network to become complex with longer pipelines. Therefore, line pressure, deterioration, aging, natural disasters and man-made destruction [1] are threatening the safety and efficiency of gas pipeline. Research in the area of pipeline leakage detection and localization has begun from 1970s and some results have been applied to industry which developed many new techniques and methods which have obtained many achievements and are applied to practice. At present, the methods of pipeline leak detection and location [2] are mainly divided into two parts: direct detection method and indirect method for detecting [3].

In this paper, taking Daqing natural gas pipeline leak detection technology applied research as the background, we make a de-noising preprocess through the wavelet. Then we use RBF neural network algorithm to determine whether the signal denoised is a leakage signal or not. Process the leakage signal by DS evidence theory data fusion algorithm comprehensively, we can determine the exact location of the leak.

2. Signal Preprocessing based on Wavelet Transform

Fourier transform theory has been widely used in many fields of science and technology. The advantage of the Fourier analysis is the ability to clearly show that signal frequency characteristics, however, it can't reflect the local information in the time domain. Local qualitative description is very important either in theoretical

research or in practical application. In order to get rid of the limitations of Fourier analysis, the researchers has developed and improved Fourier transform from different aspects. Up to now, the wavelet transform in numerous improvements is the most profound mathematical analysis method. The main idea of the short-time Fourier transform [3] is: the signal is split into many small time blocks, using Fourier transform analyze every block of time to determine the frequency in the block.

Fourier further proposed the function that, the cycle is 2π can be represented by a series of trigonometric function value, namely:

$$f(x) \approx \frac{a_0}{2} + \sum_{k=1}^{+\infty} [a_k \cos kx + b_k \sin kx] \quad (1)$$

where:

$$a_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos kx dx, \quad b_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin kx dx.$$

Expression (1) can be treated as signal $f(x)$ is composed by multiple sine wave superposition, among them, a_k, b_k is overlay weight, said the signal in different frequency amplitude of the size.

Obviously, when the signals is the symmetry signals:

$$b_k = 0, \quad f(x) \approx \frac{a_0}{2} + \sum_{k=1}^{+\infty} a_k \cos kx \quad (2)$$

When the signals is the antisymmetric signals:

$$a_k = 0, \quad f(x) \approx \frac{a_0}{2} + \sum_{k=1}^{+\infty} b_k \sin kx \quad (3)$$

In the process of research on heat conduction equation, Fourier transferred heat conduction equation from time domain to frequency domain to simplify the original problem, hence, the Fourier transform was proposed.

Assume signal $f(x)$ has interest in the frequency around the time $x = \tau$, obviously, only get the value of FFT at the time I_τ is a simple method, namely:

$$\hat{f}(\omega, \tau) = \frac{1}{|I_\tau|} \int_{I_\tau} f(x) e^{-ix\omega} dx \quad (4)$$

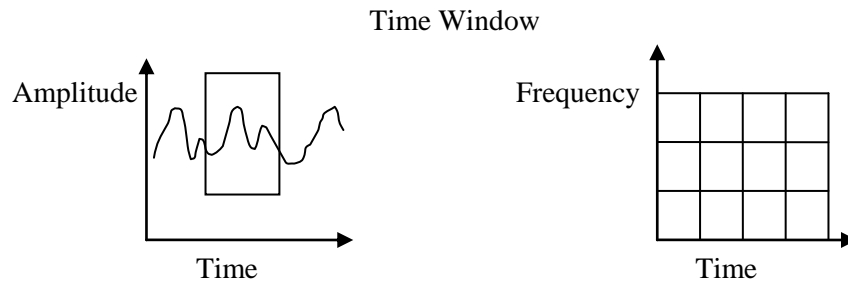
where $|I_\tau|$ represent the length of the area I_τ . If the square wave $g_\tau(x)$ is defined as:

$$g_\tau(x) = \begin{cases} \frac{1}{|I_\tau|}, & x \in I_\tau \\ 0, & \text{others} \end{cases} \quad (5)$$

Then formula (4) is:

$$\hat{f}(\omega, \tau) = \int_{\mathbb{R}} f(x) g_\tau(x) e^{-ix\omega} dx \quad (6)$$

where \mathbb{R} is the whole real axis.



(a) Time-domain add window (b) Time-frequency plane partition

Figure 1. STFT Schematic Diagram

In the threshold de-noising, threshold function for below threshold and higher threshold of wavelet system mathematical model, there has different estimation methods and different processing strategies. Set ω is the original wavelet coefficients, $\eta(\omega)$ is wavelet coefficients after a threshold, and T is threshold.

$$I(x) = \begin{cases} 1, & x \text{ True} \\ 0, & x \text{ False} \end{cases} \quad (7)$$

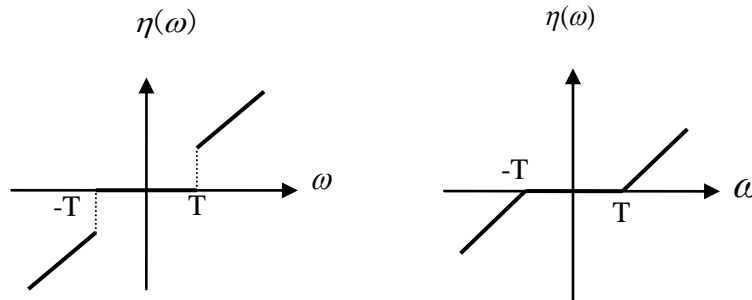
There is the common threshold function [5] which has two following kinds: hard threshold function and soft threshold function.

Hard threshold function, as shown in figure 2(a):

$$\eta(\omega) = \omega I(|\omega| > T) \quad (8)$$

Soft threshold function, as shown in figure 2(b):

$$\eta(\omega) = (\omega - \text{sgn}(\omega)T) I(|\omega| > T) \quad (9)$$



(a) Hard threshold function

(b) Soft threshold function

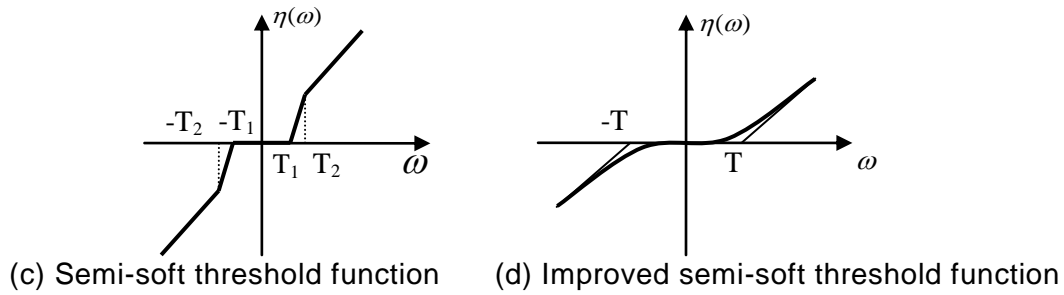


Figure 2. Threshold Function

In the Figure 2, ordinate is the wavelet coefficients after a threshold, and abscissa is the original wavelet coefficients of signal. Analyze the hard threshold and soft threshold method under the condition of Gaussian noise produced by the deviation, variance and L_2 risk formula, could get the following several conclusions:

(1) Give the threshold T , Hard threshold always have a bigger variance than soft threshold.

(2) When coefficient is big enough, the deviation is smaller that caused by the hard threshold variance than the soft threshold variance caused.

(3) When coefficient is nearby T , the hard threshold method has a greater deviation, L_2 risk and variance. When coefficient is small, the L_2 risks of two methods are quite small.

Although the soft threshold processing is relatively smooth, but may produce edge blur distortion, *etc.* However hard threshold processing is a good way to retain local characteristics of signal edge, but there will be some distortion such as pseudo Gibbs effect and ringing.

In order to avoid the defect of hard threshold method, a semi-soft threshold function was proposed, as shown in Figure 2(c), the expression is as follows:

$$\eta(\omega) = \text{sgn}(\omega) \frac{T_2(|\omega| - T_1)}{T_2 - T_1} I(T_1 < |\omega| < T_2) + \omega I(|\omega| > T_2) \quad (10)$$

Where, $0 < T_1 < T_2$.

Semi-soft threshold method needs to estimate two different thresholds, in order to make the soft threshold method with a higher order. The improved expression is as follows:

$$\eta(\omega) = \begin{cases} \omega + T - \frac{T}{2k+1}, & \omega < -T \\ \frac{1}{(2k+1)T^{2k}} \omega^{2k+1}, & |\omega| \leq T \\ \omega - T + \frac{T}{2k+1}, & \omega > T \end{cases} \quad (11)$$

From Figure 2(d) we can see the improved soft threshold function has a smooth transition region between the useful signal and wavelet coefficient of the noise.

3. Leakage Signal Identification based on RBF Neural Network

With the improvement and development of the artificial neural network theory and technology, the traditional way of thinking was broken, and it becomes a new way to solve complex problems. Introducing artificial neural network in the natural gas pipeline leak detection system has more efficient, more stable, and more accurate. RBF neural network is introduced in this chapter to construct the model of natural gas pipeline judge leakage signal, it is used to comprehensive judge the signal which have went through the wavelet de-noising processing whether is the leakage signal

3.1. Artificial Neural Network

Artificial neural network [6] is a kind of mathematical model which is distributed, parallel processing and to imitate the neural network of animal behavior characteristic. This type of network by adjusting its internal relationship between a large numbers of nodes, could achieve the destination of information processing.

Neural network has the following two relatively significant computing capabilities: (1) the larger scale of parallel and distributed structure. (2) Neural network has strong ability of self-learning and generalization ability. Generalization is pointed to if the input data is not learning (training), the output of the neural network also can get reasonable. Then neural networks can make use of the information processing's abilities to solve some complex and large problems.

But in the practical applications, in a systems engineering, the neural network needs to be combined with other methods, it cannot get answer as a separate method. Specifically, some complex problems should be broken down into several relatively easy problems, and then can use neural networks to solve one or a few sub-problems that within its ability

Its basic structure is neurons, which can deal with information, Figure 3 is the basic neuronal model, and it is a well-designed foundation of artificial neural network. A neuron has three basic elements:

(1) Synaptic also called link chain, the characteristics of each synapse usually represented as weight or strength. In particular, the input signal x_j of the synapse j which connects to the neuronal K multiplies the synaptic weight ω_{kj} , the K is query neurons, and the j is the neuron weights of neuron synaptic input end which are the subscripts of ω_{kj} . The neuron synaptic weights can get positive or negative value, its value in the range, and the synapses of the brain is not the same.

(2) Adder, be used to calculate the sum of input signal after neurons synaptic corresponding weighted. It actually could be seen as a linear combination.

(3) Activation function that is used to limit the output amplitude of the neurons. Activation function sometimes also be refereed as suppression function, it is because sometimes that it will limit the value of the output signal in a certain permission scope. In general, a neuron output should through the normalized processing, the scope of its normal range in a closed interval $[0,1]$ or another range $[-1,+1]$.

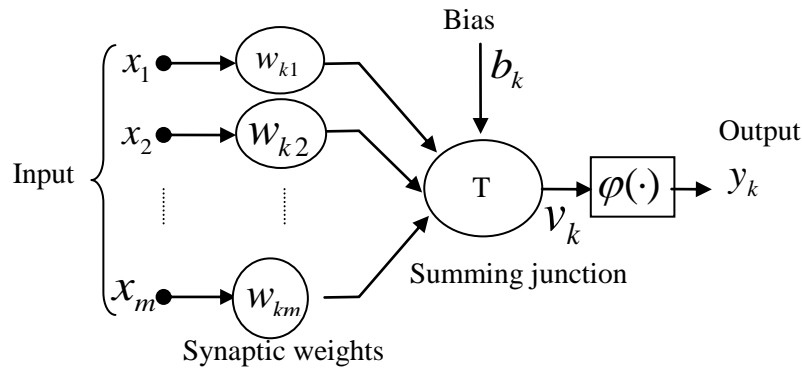


Figure 3. Nonlinear Model of Neurons

3.2. Radial-Basis Function Network

Radial-Basis Function (RBF) method with Multivariable Interpolation which is also known as the RBF neural network, structurally, RBF constructed by neural network is a multilayer feed forward neural network.

From the structure, the RBF neural network is a multilayer feed forward neural network. It contains input layer, hidden layer and output layer that consist of a three-layer feed-forward network. Signal source node composition the input layer of the network. The second layer is hidden layer, according to the problems described to determine the number of hidden layer nodes, the nodes of transformation function that it is attenuation, nonnegative nonlinear and center of radial symmetry. The third layer is output layer of input mode to respond.

The construction of RBF neural network's basic idea is: hidden layer nodes of hidden layers in the space of the "base" using radial basis function (RBF), input vector through the hidden layer to transform. Constitute a space in the hidden layer nodes of hidden layers on the "base" using radial basis function (RBF), input vector through the hidden layer to transform. Use of high dimension space replaces low dimensional space to transform the data of model input, the advantage of this is high dimension space can liner divide and low dimension space can't liner divide problems.

There are M neurons in the input layer of RBF neural network model, There are $I(I < N)$ neurons in hidden layer, an arbitrary neuron is defined as i , the excitation output of i hidden layer node is called as $\phi(X, t_i)$, in which $t_i = [t_{i1}, t_{i2}, \dots, t_{im}, \dots, t_{iM}]$ ($i=1, 2, \dots, I$) is the center of basis function. Output layer has J neurons in which an arbitrary neuron expresses j , the synaptic weights are expressed by using w_{ij} ($i=1, 2, \dots, M, j=1, 2, \dots, P$) between output layer and hidden layer.

3.3. The Signal Identifying Type based on RBF Neural Network

In general, the more training data, the more training result reflect its internal rules accurately, but there is too much sample data, it will be difficult to improve the precision of neural network. Because of RBF neural network uses the "8-4-1" structure in this paper, chooses a monitoring data as training samples. The data was normalized after pretreatment, neural network was trained, then we could get the error curve of network training as shown in Figure 4.

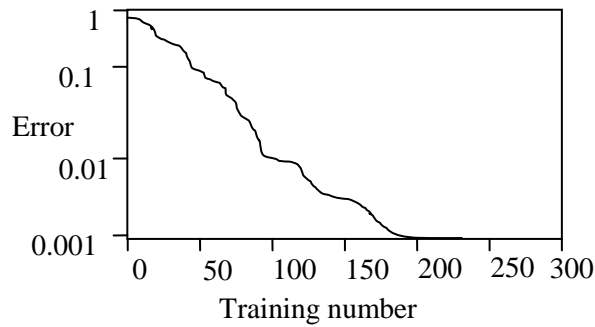


Figure 4. The Training Error Curve of RBF Neural Network

From Figure 4, we can see the training error of the the RBF neural network is larger in the beginning of, with the increase of training frequency, error is smaller and smaller. When error achieves an acceptable range, it can stop the RBF neural network training.

We choose the two groups of the actual parameters when the gas pipeline normal operation and leakage as the test sample testing after training of RBF neural network, the test results as shown in Table I.

Table I. Test Sample Output

Test Sample	RBF Neural Network Output	Ideal Output
Normal	0.10	0
Normal	0.06	0
Leakage	0.91	1
Leakage	0.85	1

From Table I ,we can see that after training, the RBF neural network testing results close to the ideal value and diagnosis structure more accurately.

4. Leak Detection based on multi-sensor data fusion

Multi-source data fusion (MSDF) [7-8] also known as information fusion, is to simulate the human brain processing of dealing with some complex problems comprehensively [9]. Data fusion can make multi-information sources work collaboratively, so as to improve the effectiveness of system. Figure 5 is multiple sensor data fusion processing model.

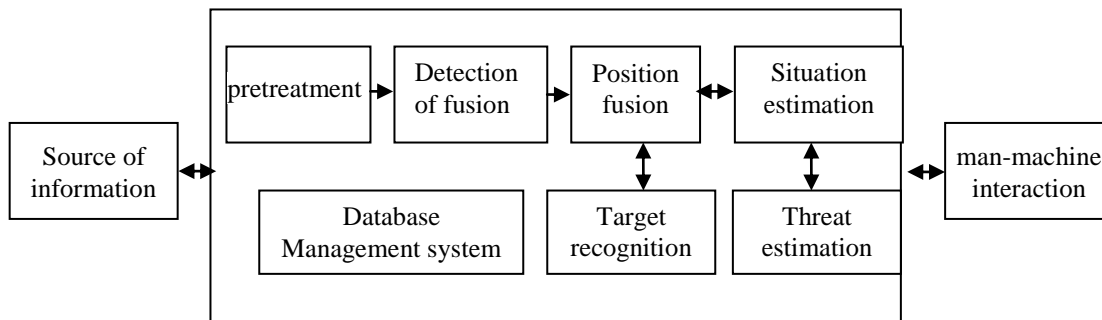


Figure 5. Multi-sensor Data Fusion Processing Model

4.1. The Combination Rule of Evidence Theory

Evidence theory, which is proposed by Dempster, and then Shafer enlarged and promoted of, so it is also known as D-S evidence theory [10]. Then the combination rule of evidence theory was put forward, suppose m_1, \dots, m_n is the basic probability assignment of n independence on 2^U , BEL_1, \dots, BEL_n is n belief functions on the same recognition frame U, m_1, \dots, m_n is the basic probability value respectively, and set:

$$K_1 = \sum_{A_i \cap B_j \cap \dots \cap Z_k \neq \phi} m_1(A_i) \cdot m_2(B_j) \cdot \dots \cdot m_n(Z_k) < 1 \quad (12)$$

Then, we can get:

$$m(C) = \begin{cases} \frac{\sum_{A_i \cap B_j \cap \dots \cap Z_k = C} m_1(A_i) \cdot m_2(B_j) \cdot \dots \cdot m_n(Z_k)}{1 - K_1}, & \forall C \subset U, C \neq \phi \\ 0, & C = \phi \end{cases} \quad (13)$$

In the formula, if $K_1 \neq 1$, m defines the basic probability value; If $K_1 = 1$, m_1, \dots, m_n will be contradicted, there are no associated basic probability value.

4.2. The Recursive Target Identification Fusion of Incompatible Data Structure

Assumes the accumulation of uncertainty that assigned to the recognition framework $\theta^i(k-1) = 1 - \sum_{j=1}^N m_j^i(k-1)$ and the accumulation of basic probability assignment $m_j^i(k-1), i=1, \dots, M; j=1, \dots, N$ to determine the i th sensor at the moment (K - 1) that about the incompatible and accumulation information of target recognition. At the moment of k , sensors i can get the measurement uncertainty of object recognition $\theta_k^i = 1 - \sum_{j=1}^N m_{jk}^i$ and the new measurement basic probability assignment of target recognition m_{jk}^i . According to the rules of Dempster combination, we can obtain the k times of the accumulation uncertainty $\theta^i(k) = 1 - \sum_{j=1}^N m_j^i(k)$ and the accumulation basic probability assignment $m_j^i(k)$ of target recognition from Table II.

Table II. Time Domain Identification Evidence Combination of Separation Structure

	m_{1k}^i	m_{2k}^i	...	m_{Nk}^i	θ_k^i
$m_1^i(k-1)$	$m_1^i(k-1)m_{1k}^i$	$m_1^i(k-1)m_{2k}^i$...	$m_1^i(k-1)m_{Nk}^i$	$m_1^i(k-1)\theta_k^i$
$m_2^i(k-1)$	$m_2^i(k-1)m_{1k}^i$	$m_2^i(k-1)m_{2k}^i$...	$m_2^i(k-1)m_{Nk}^i$	$m_2^i(k-1)\theta_k^i$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
$m_N^i(k-1)$	$m_N^i(k-1)m_{1k}^i$	$m_N^i(k-1)m_{2k}^i$...	$m_N^i(k-1)m_{Nk}^i$	$m_N^i(k-1)\theta_k^i$
$\theta^i(k-1)$	$\theta^i(k-1)m_{1k}^i$	$\theta^i(k-1)m_{2k}^i$...	$\theta^i(k-1)m_{Nk}^i$	$\theta^i(k-1)\theta_k^i$

Because elements are incompatible, so $\forall l \neq j, m_j^i(k-1)m_{jk}^i$ is conflict information, Through the Dempster combination rules, the new Accumulation of basic probability assignment $m_j^i(k), i=1, M, j=1, N$ can be according to the following formula:

$$m_j^i(k) = \frac{m_j^i(k-1)m_{jk}^i + m_j^i(k-1)\theta_j^i + \theta^i(k-1)m_{jk}^i}{1 - K_k^i} \quad (14)$$

Among them:

$$K_k^i = \sum_{j \neq l} m_{lk}^i m_j^i(k-1) \quad (15)$$

The accumulation uncertainty of update is:

$$\theta^i(k) = \frac{\theta^i(k-1)\theta_k^i}{1 - K_k^i} \quad (16)$$

By repeating the above process can obtain the results of sensors in time domain for the target identification fusion.

When all the sensors get the cumulative uncertainty of target recognition by the recursive way in the time domain as well as accumulate all the basic probability values, time domain cumulative information of sensors M uses spatial domain fusion by using the Dempster combination rule [11, 12]. We can get sensors i and l's final fusion results of the time/space accumulative target identification by using the space information fusion of different sensors, which is shown as following:

$$m^{il}(k) = \frac{m_j^i(k)m_j^l(k) + m_j^i(k)\theta^l(k) + \theta^i(k)m_j^l(k)}{1 - K_k^{il}} \quad (17)$$

In the formula:

$$K_k^{il} = \sum_{j \neq n} m_j^i(k)m_n^l(k)$$

The accumulative time/space uncertainty which gets from sensors i and l is as following:

$$\theta^{il}(k) = \frac{\theta^i(k)\theta^l(k)}{1 - K_k^{il}} \quad (18)$$

In the incompatible data structure condition, we obtain a recursive temporal-spatial fusion model based on target recognition of D-S evidence theory. Repeating the above process, we can get the result that mixed M sensors the total accumulation of time and space target identification information.

Table III. Space Target Recognition Information Fusion of Different Sensors

	$m_1^i(k)$	$m_2^i(k)$...	$m_N^i(k)$	$\theta^i(k)$
$m_1^i(k)$	$m_1^i(k)m_1^l(k)$	$m_1^i(k)m_2^l(k)$...	$m_1^i(k)m_N^l(k)$	$m_1^i(k)\theta^l(k)$
$m_2^i(k)$	$m_2^i(k)m_1^l(k)$	$m_2^i(k)m_2^l(k)$...	$m_2^i(k)m_N^l(k)$	$m_2^i(k)\theta^l(k)$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
$m_N^i(k)$	$m_N^i(k)m_1^l(k)$	$m_N^i(k)m_2^l(k)$...	$m_N^i(k)m_N^l(k)$	$m_N^i(k)\theta^l(k)$
$\theta^i(k)$	$\theta^i(k)m_1^l(k)$	$\theta^i(k)m_2^l(k)$...	$\theta^i(k)m_N^l(k)$	$\theta^i(k)\theta^l(k)$

4.3 Experimental Simulation

Suppose $U = \{A_1, A_2\}$ for recognition framework, A_1 said “normal operation”, A_2 said “having leak”. $m_i(A_j)$ ($i=1,2,\dots,7,8$)($j=1,2$) said the basic probability assignment of 8 sensors, Respectively is the upstream Pipe internally pressure, the upstream instantaneous flow, the upstream acoustic monitoring data, the upstream Negative pressure wave monitoring data, the downstream Pipe internally pressure, the downstream instantaneous flow, the downstream acoustic monitoring data, the downstream Negative pressure wave monitoring data, the upstream is DaQing Honggang station and the downstream is DaQing Xingsan Dingzikou station, their distance is 3.13Km. $m_i(U)$ ($i=1,2,\dots,7,8$) is unknown probability. In general, it can be obtained $m_i(A_j)$ by evidence synthesis formula. In this paper, the 8 sensors had been imported into “8-4-1” structure of RBF neural network, the probability of each sensor after neural network distribution as shown in Table II.

The D-S evidence synthesis rules, synthetic result and leakage location as shown in the table.

The Table IV shows that D-S evidence theory when used to judge the leakage position is relatively ideal, and more accuracy. The error is less than the line length $\times (\pm 0.5\%) \pm 50m$.

Table IV. D-S Evidence Theory Results and Leak Location

Date Probability	2011.11.12	2011.11.13	2011.11.14	2011.11.15	2011.11.16	2011.11.17	2011.11.18
$m(A_1)$	0.9391	0.9758	0.8645	0.0461	0.0416	0.0081	0.0984
$m(A_2)$	0.0606	0.0234	0.1349	0.9530	0.9576	0.9917	0.9011
$m(U)$	0.0003	0.0008	0.0006	0.0009	0.0008	0.0002	0.0005
Leak location	No Leaks	No Leaks	No Leaks	From Honggang 1.24km	From Honggang 1.36km	From Honggang 1.27km	From Honggang 1.34km
actual leak location	No Leaks	No Leaks	No Leaks	From Honggang 1.30km	From Honggang 1.30km	From Honggang 1.30km	From Honggang 1.30km

5. Conclusion

First of all, in this paper, the sound waves and negative pressure wave signals which are uploaded from the lower machine sensor are conducted noise filtering by using soft threshold function and SURE Shrink threshold estimation, we can obtain a better denoising effect and avoid to preprocess the original signal before the RBF neural network and data fusion; then according to actual demand to construct the "8-4-1" RBF neural network model which has three layers, the signal after denoising process, coming from eight sensors, represents the input parameter of RBF neural network model. Based on the RBF neural network, we can determine the output signal is a leakage signal or not. If it is the leakage signal, the signal could be treated as the input parameters of data fusion to establish the data fusion model based on D-S evidence theory for solving the localization problems of detecting the pipeline leakage.

References

- [1] M. Taghvaei, S. B. M. Beck and M. J. Staszewski, "Leak detection in pipelines using cepstrum analysis", *Measurement Science and Technology*, vol. 17, (2006), pp.367-372.
- [2] M. Taghvaei, S. B. M. Beck and M. J. Staszewski, "Leak detection in pipelines using cepstrum analysis", *Measurement Science and Technology*, vol. 17, (2006), pp. 367-372.
- [3] X. Wang, K. Huang and X. Zhu, "Oil and gas pipeline leak detection technology", *Pipeline technology and equipment*, vol. 1, (2008), pp. 24-25.
- [4] L. Cheng and Hanwei Guo, "The theory and engineering practice of wavelet and discrete transform", Beijing: Tsinghua University press, (2011), pp. 1-9.
- [5] H. Fei, X. Shi, X. Zhao and L. Yang, "Wavelet denoising algorithm Based on improved threshold function", *Modern electronic technique*, vol. 34, no. 12, (2011), pp. 61-64.
- [6] J. Tian and M. Gao, "Research and applications of artificial neural networks", Beijing: Beijing Institute of Technology press, (2006), pp. 35-40.
- [7] Y. He, G. Hong Wang and X. Guan, "Information fusion theory and application", Beijing: Electronic industry press, (2010), pp. 2-7.
- [8] L. He, G. H. Huang and G. M. Zeng, "Wavelet-based multiresolution analysis for data cleaning and its application to water quality management systems", *Expert Systems with Applications*, vol. 3, no. 35, (2008), pp. 1301-1310.
- [9] Y. He, G. Wang and X. Guan, "Information fusion theory and application", Beijing: Electronics industry press, (2010), pp. 2-7.
- [10] B. Chen, "Multi-source data fusion method and key technology research of gas pipeline leakage monitoring network", Beijing University of Posts and Telecommunications, (2009).
- [11] B. Chen, J. W. Wan, Y. F. Wu and N. Qin, "Neural networks and evidence theory fusion method of pipeline leakage diagnosis", *Journal of Beijing University of Posts and Telecommunications*, vol. 3, no. 1, (2009), pp. 5-9.
- [12] S. Fei Jiang, C. Mei Zhang and ZiWei Jin, "Damage Identification Method Based on Back-Propagation Neural Network and D-S Evidential Theory", *Journal of Shenyang Jianzhu University*, vol. 23, no. 1, (2007), pp. 1-5.

Authors



Bingkun Gao, he received M.S degree and Doctor Degree of Engineering in control theory and control engineering from Harbin Engineering University, People Republic of China in 1995 and 2003, and stood out from Northeastern University as post doctor in Control Science and Engineering. He is currently the dean of Electrical and information engineering institute in Northeast Petroleum University. His currently research interests include Power system transmission control and fault diagnosis, multi-resource network monitoring and System simulation.

