Application of a Wavelet Packet and SOM Neural Network in Wastewater Treatment Fault

Ming Lu
Department of Mechanical and Electrical Engineering, Chizhou Vocational and Technical College, Anhui 247000, China
275683959@qq.com

Abstract. In order to solve the problem of fault diagnosis in sewage treatment, a fault diagnosis method combining wavelet packet and improved self-organization mapping (SOM) neural network was proposed. In this method, the fault of sewage treatment is decomposed into three types by wavelet packet, and then decomposed into several frequency bands to calculate the energy of different frequency bands. By using the ratio of these energy values to the energy values of normal operating frequency bands, the feature vector of fault diagnosis of sewage treatment fault is constructed to extract fault features. SOM neural network of 3 x 3 was designed, and the neural network training was carried out using the energy fault feature vector [1], so as to determine the network parameters and achieve the purpose of fault diagnosis. The simulation results show that the fault diagnosis method is effective and accurate.

1. Introduction
In the process of water pollution treatment which characteristics of sewage treatment are strong coupling, non-linearity and strong hysteresis, which will cause various failures during long-term operation. For example, it consider to a number of factors: weather and water inflow, flow rate and temperature, this will lead to the deviation of the water parameters of the key parameters BOD and COD, thus impacting the activated sludge. In the water treatment process, it is divided into the following parts: pretreatment, primary sedimentation, aeration, secondary sedimentation and sludge reflux, etc., and in these processes, there are abnormal pH changes, insufficient supply, sludge increase and Changes in key parameters of water quantity. These conditions will lead to the failure of sewage treatment equipment. Secondly, due to the longer running time of the equipment used, as the equipment ages and corrodes, some sensors may be damaged, resulting in errors of important parameters BOD and COD. In the long-term use of these equipment, some key equipment such as heat exchangers, circulation devices, and blowers will also malfunction, timely detection of faults and effective diagnosis of faults, have an important role in sewage treatment. At present, the main methods of fault diagnosis fall into two categories: qualitative analysis methods and quantitative analysis methods. Due to the complex nature of wastewater treatment, traditional methods cannot be accurately judged and diagnosed. In recent years, with the rapid development of artificial intelligence, there are many ways to deal with faults: fault tree, expert system, neural network, principal component analysis, support vector machine, spectral analysis, wavelet analysis, etc. Get the app. For example, Heikkinen proposed the SOM neural network analysis process for wastewater treatment. The secondary method is applied to the som neural network by using k-means
clustering, which effectively classifies the data and ensures that a certain state corresponds to a certain kind. The degree of change in a particular state. [1]YU proposed the SOM model for PCA and redundant analysis optimization. This model is applied to predict dissolved oxygen and ammonia nitrogen, and the prediction effect is good [2].

In this paper, the three-layer wavelet packet theory is used to decompose, and the energy of each frequency band in the third layer is extracted and normalized, and then input as the characteristic value of the fault mode. The input layer of the self-organizing neural network SOM is used as the feature vector of the fault mode. According to the characteristics of the self-organizing neural network SOM, it has good self-organization, self-adaptation and robustness. It can learn and simulate the unknown environment. It also makes appropriate adjustments to its own network structure and is able to extract features of the input signal pattern. This utilizes the energy of each node to reflect the characteristics of the fault signal and is convenient to apply. At the same time, feature values can be extracted from the collected data, and the data has a good classification effect. But the disadvantage is that the som network relies on the weight, combined with the wavelet packet theory, to judge the fault, the effect is obvious.

2. Wavelet packet theory and feature extraction

2.1. Wavelet packet decomposition

Wavelet packet theory was introduced in 1989 by the founders of wavelets Meyer, Coifman and Wickerhauser. [3]The wavelet packet is an effective analysis method for the signal. The main purpose is to divide the frequency band into multiple layers, effectively solving the shortcomings of the wavelet with poor frequency resolution in the high frequency band, and then selecting the corresponding frequency band according to the characteristics of the signal. The corresponding signal spectrum is matched to improve the time-frequency resolution [4].

For example, given a specific signal, the signal is decomposed into various frequency bands through a set of orthogonal high and low filters H, G, taking a three-layer wavelet decomposition as an example, where H represents a low frequency and G represents a high frequency [4]. As shown below:

![Wavelet decomposition process](image)

Therefore, based on the decomposition characteristics of the wavelet packet, the fault signal is set to X (t), and the algorithm for decomposing X (t) is

\[
\begin{align*}
    x(t) &= s_0^1(t) \\
    s_i^{2i-1}(t) &= \sum_m h(k-2t)s_{i-1}^1(t) \\
    s_i^{2i}(t) &= \sum_m g(k-2t)s_{i-1}^1(t)
\end{align*}
\]  

(1)

Where j is the number of layers in the wavelet packet tree structure, i is the band node, i=1, 2, •••, n, n is the number of j-layer wavelet packet points, k is the translation factor, and h (k) is the low-frequency filter, g (k) = (-1) ^k h(1-k) is a high-pass filter with orthogonality [5].
2.2. **Wavelet packet energy characteristic extraction process**

In the time domain transformation process, the wavelet transform satisfies its energy conservation theory:

\[
\int_{-\infty}^{\infty} x(t)^2 dt = \sum_{i=1}^{n} s_i^j(t)^2, \tag{2}
\]

\(s_i^j(t)\) Indicates the energy level of each band. Then, the energy represented by the reconstructed signal \(x(t)\) of the \(i\)-th band of the \(j\)th layer \(D_i^j\):

\[
D_i^j = \sum_{i=1}^{n} s_i^j(t)^2 \tag{3}
\]

Then the \(i\)-th band energy normalization process of the \(j\)-th layer can be expressed as:

\[
E_i^j = D_i^j / \sum_{i=1}^{2j-1} D_i^j \tag{4}
\]

\(E_i^j\) Is the energy of the \(i\)-th band of the \(j\)th layer. Faults in different situations have different signal representations. To detect and quantify these modes, the signal energy characteristics can be obtained by calculating the energy and normalization of signals in different frequency bands for different signal energy at different wavelet packet nodes [6]. The energy distribution of each frequency band and the signals in different states have different energy distributions of the frequency band. In this paper, the original signal is decomposed into three layers by wavelet packet transform, and the energy of each frequency segment of the third layer is calculated separately, and the energy values and normal working hours are utilized [7]. The ratio of the energy values of each frequency band can construct the characteristic vector of the sewage treatment failure mode discrimination:

\[
T = [E_1^1, E_2^1, E_3^1, \ldots, E_{2j-1}^1] \tag{5}
\]

3. **SOM neural network learning algorithm**

3.1. **Principle of SOM Algorithm**

The self-organizing neural network SOM adopts a non-tutor learning algorithm. According to its own network characteristics and learning mode, through the repeated learning of the input mode, the basic characteristics of each mode are obtained, and self-organization is carried out in the competition layer [8]. The winning neurons show that SOM can learn and simulate the unknown environment, and can also adjust its network characteristics. The SOM neural network is further divided into processing unit, network structure and learning rules, and is composed of two layers, which are input and output layers respectively. The processing unit functions to simulate biological neuron functions, and each processing unit has multiple inputs and multiple An output path is used for the transmission of information, and the result of the information processing is output to the outside world. The input layer neurons transmit information to the output layer neurons through the weight vector, and there is no connection between the output layer and the input layer. Because the output layer exhibits competitive characteristics, the output layer is the competition layer. Nodes with different output layers represent different classification modes. As shown below:
As shown in Figure 2, the network has a total of \( M \times M \) neurons arranged into a two-dimensional matrix, the input neurons are in the lower layer, and there are \( N \) neurons. All the input layer neurons are used between the output layer neurons and the right. The process of connecting, and there are also local links between the output layers. The SOM neural network trains the network unsupervised and classifies the input patterns.

The SOM learning algorithm is divided into three processes of competition, cooperation and update.

1. Set the input variable \( X = [x_1, x_2, \cdots, x_N] \). Weight vector \( W = [w_{i1}, w_{i2}, \cdots, w_{iN}]^T \).

   Neuron input
   \[
   u^i = \sum_{j=1}^{N} w_{ij} x_j
   \]  

2. Calculate the Euclidean distance between the input vector and the connection weight vector, determine the neuron with the smallest Euclidean distance, and record the winning neuron as \( A \). Wherein, the Euclidean distance between the \( i \)-th input vector and the \( j \)-th neuron:
   \[
   d_j = \text{MIN}(x - w_i)
   \]  

The SOM learning algorithm is as follows:

1. According to the calculation formula of the Euclidean distance, the minimum intra-body neuron \( J^* \) is calculated, that is, a certain unit \( K \) is determined:
   \[
   d_k = \text{min}(d_j)
   \]  

2. Give a surrounding field \( s(t) \): modify the weight of the output neuron \( J^* \) and the leading neuron according to the following formula:
   \[
   w_{ij}(t + 1) = w_{ij}(t) + \rho(t)[x_i(t) - w_{ij}(t)]
   \]  

   In the above formula, \( \rho \) is a gain term and falls with time until 0 is usually taken:
   \[
   \rho(t) = \frac{1}{t}
   \]  

3. Calculate the output \( O_k \):
   \[
   O_k = f\{ \text{MIN}(x - w_i) \}
   \]  

4. Update the learning rate \( \rho(t) \) and the topological neighborhood \( s(t) \), and renormalize the learned weight vector.
(5) Determine whether the end condition is satisfied $\rho(t) \leq \rho_{\text{min}}$, if it is satisfied, the training is ended; if it is not satisfied, it is judged whether the maximum number of iterations is reached, and if the maximum number of iterations is not reached, the iteration is continued, otherwise the training is ended.

4. Fault diagnosis experiment and result analysis

4.1. Fault sample database
During the sewage treatment process, the common five failure modes of the value of the output value in the normal state and the value in the abnormal state are counted. And in the initial signal to remove the influence of redundant signals and noise signals, the samples of the commonly used five failure modes can be obtained by simulation results. These fault samples are decomposed as wavelet packets. If the wavelet packet is divided into three layers, that is, $2^3=8$ frequencies. The wavelet packet base (no overlap), the result of using 8 wavelet bases corresponds to the corresponding frequency band, and the SOM neural network input is the energy vector ratio between the value of the corresponding frequency band and the normal value as the input feature vector. However, due to the small number of samples obtained, the first 30 sets of data can be used as training samples, and the latter 2 sets are used as test samples.

In this example, the set of five fault samples are: anoxic bacteria fault P1, autotrophic bacteria fault P2, dissolved oxygen sensor fault P3, nitrate nitrogen fault P4, equipment instrument fault P5. SOM network for fault diagnosis, sample the following table:

| cause of issue | p1  | p2  | p3  | p4   | p5   |
|----------------|-----|-----|-----|------|------|
| t1             | 0.2584 | 0.8432 | -0.632 | 0.974 | 0.632 |
| t2             | 0.23  | 0.5251 | -0.1  | 0.563 | 0.87  |
| t3             | 0.874 | 0.42  | -0.5  | 0.432 | 0.431 |
| t4             | 0.429 | -0.4  | -0.34 | 0.65  | 0.674 |
| t5             | 0.726 | -0.45 | -0.9  | 0.43  | 0.54  |

4.2. SOM Network Training
For the description of the above fault categories and the required training samples, the competition layer of the SOM network can be selected as a $3 \times 3$ array of neurons, a hexagonal topology (hextop), and the number of training times of the input vector is set to 250, the neighborhood size. The initial value is 3, and the neuron distance uses the linkdist function to calculate the distance. To prevent deviations between the input sample data, the data is first normalized when the network is created. Figure 3 shows the training process records of the SOM neural network. The weight distances of adjacent neurons obtained after network training are shown in Figure 4. The number of sample points on each neuron is shown in Figure 5.
As shown in the above figure, the gray square indicates the neurons, and the line between the gray squares represents the connection between the neurons. The distance between the neurons can be calculated by the Euler formula to calculate the nerve. The Meta distance is also the distance between the squares. The background color of the hexagon of the neuron shown in Figure 4 reflects the distance between the neurons. If the color is darker, it means that the distance between the neurons is farther, and vice versa. The difference in the corresponding state of the two neurons is determined by the distance between the neurons. The farther the distance is, the larger the difference is, and the closer the difference is. Figure 4 and Figure 5 show that the position of the output neuron is the same or similar to the position of the competing victorian neuron corresponding to a standard sample, indicating that a fault has occurred in the detected sample; that is, the position of the output neuron in the output layer is in many standard states. Between, it indicates that there may be several different state faults, and the degree of fault is determined by the Euclidean distance between the location and the corresponding standard location.

In order to detect the network's judgment of the fault diagnosis effect, the five sets of data shown in Table 1 were tested. These sample data are first imported into the trained SOM neural network, as shown in the matrix information below. Look at the output corresponding to each set of data. The network
output corresponding to the data samples of Groups 1 and 2 is 6, the network output corresponding to the 3rd and 4th samples is 1, and the network output of the 5th and 6th samples is 8, the 7th and 8th samples. The corresponding network output is 3, that is to say, through the classification of the network, the samples in the first and second groups are faults of the oxygen-producing bacteria, the samples in the third and fourth groups are the faults of the autotrophic bacteria, and the samples in the fifth group are the dissolved oxygen. For sensor failure, the sixth group of samples is nitrate-nitrogen failure, and the seventh and eighth groups of samples are fault-free, which can lead to correct conclusions and accurate methods.

\[
y = \begin{bmatrix}
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 
\end{bmatrix}
\]

4.3. Experimental analysis
(1) In the process of wavelet packet and SOM network analysis failure, the effect of selecting wavelet function has great influence on the result of fault analysis. It may happen that because of the large number of wavelet functions, the accuracy of fault diagnosis is 70%. Even higher.
(2) Adding a white noise signal in the original stage of the data signal may cause some kind of fault to be inaccurate, so there is an error when sorting with Som.

4.4. Deficiency and improvement
The shortcomings of the SOM algorithm are obvious. First, in the classification process, the learning rate and stability cannot be unified, and the weight dependence is serious. Second, if a neuron deviates far from other neurons, it becomes impossible to learn. Dead neurons.
Improvement: Usually, some intelligent algorithms, such as particle swarm optimization, imperial competition algorithm, etc., can be used to optimize weights.

5. Conclusion
(1) This paper puts forward the faults in sewage treatment: the problem of abnormal oxygen bacteria, autotrophic bacteria failure, dissolved oxygen sensor failure, and solution of nitrate nitrogen failure. The data of sewage treatment is obtained by matlab simulation, and the following conclusions are drawn: When using the wavelet packet to analyze the fault process, the redundant information and noise in the data should be out in time to extract the fault characteristics.
(2) The superiority of som network is better than other fault diagnosis methods. The field of application is more extensive. But how to design network parameters and weights, relying on some intelligent algorithms alone is not enough, only optimization. Therefore, we should further explore the scope of adaptation of faults and increase value.

Acknowledgments
This work was financially supported by Chizhou Vocational and Technical College fund. project number:ZR2019Z0202. Address: Room 1, Unit 1, Building 32, Hebin Garden, Guichi District, Chizhou City, Anhui Province; Tel: 13665666806
References

[1] Lihong Chen, Haomiao Xu, Jiangkun Xie, Xiaoshuang Liu, Yong Yuan, Ping Liu, Zan Qu, Naiqiang Yan. [SnS 4] 4- clusters modified MgAl-LDH composites for mercury ions removal from acid wastewater [J]. Environmental Pollution, 2019, 247.

[2] Yongyuan Hong, Guohe Huang, Chunjiang An, Pei Song, Xiaying Xin, Xiujuan Chen, Peng Zhang, Yanyun Zhao, Rubing Zheng. Enhanced nitrogen removal in the treatment of rural domestic sewage using vertical-flow multi-soil-layering systems: Experimental and modeling insights [J]. Journal of Environmental Management, 2019, 240.

[3] Hui Jiang, Ming Hua, Jin Zhang, Panpan Cheng, Zhen Ye, Mai Huang, Qiang Jin. Sustainability efficiency assessment of wastewater treatment plants in China: A data envelopment analysis based on cluster benchmarking [J]. Journal of Cleaner Production, 2019.

[4] Wu Zhenwei, Jiang Xiaoping, Ma Huimeng, Ma Suliang. Wavelet Packet for Fuzzy Energy Storage to Reduce PV Fluctuation —— Fuzzy Control [J]. Proceedings of the CSEE, 2014, 34(03): 317-324.

[5] Zhang Zhetao, Zhu Jianjun, Yan Cuilin, Zhou Wei. Wavelet Packet Multi-threshold Denoising Method and Its Application in Deformation Analysis [J]. Acta Geochimica Sinica, 2014, 43(01): 13-20.

[6] He Xing, Wang Hongli, Lu Jinghui, Jiang Wei. Fault diagnosis of analog circuits based on preferred wavelet packet and ELM [J]. Chinese Journal of Scientific Instrument, 2013, 34(11): 2614-2619.

[7] Sun Jian, Wang Chenghua, Du Qingbo. Fault diagnosis of analog circuits based on wavelet packet energy spectrum and NPE [J]. Chinese Journal of Scientific Instrument, 2013, 34(09): 2021-2027.

[8] HAN Xiaojuan, CHEN Yueyan, ZHANG Hao, CHEN Fang. Application of hybrid energy storage technology based on wavelet packet decomposition in power fluctuation of wind farms [J]. Proceedings of the CSEE, 2013, 33(19): 8-13 +24.

[9] DONG Hui. Research on Grey-RBF Neural Network Prediction Model Based on Optimized Genetic Algorithm [J]. Journal of Xinxiang University (Natural Science Edition), 2012, 29(03): 237-240.