Desalination and Water Treatment
www.deswater.com
doi: 10.5004/dwt.2023.30047

315 (2023) 637–640
December

Improvement of water pollution detection method based on convolution neural network

Weihua Liu[a], Tingting Shi[b,*], Xuetian Xu[c]

[a]*Academic Affairs Office, Guangdong Justice Police Vocational College, Guangzhou 510520, China*
[b]*College of Information Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China, email: tostthao@126.com*
[c]*Department of Information, Guangdong Justice Police Vocational College, Guangzhou 510520, China*

Received 9 December 2023; Accepted 23 August 2023

ABSTRACT

In water resources pollution detection, because the collected data is not filtered, the rate of missed detection is high, and the accuracy of detection results is limited. Therefore, a method of water resources pollution detection based on convolution neural network is proposed. The data of water resource pollution is input into convolutional neural network to calculate its average fitness value, generate the optimal rule set, calculate its average fitness, train the water resource pollution detection data, and realize the accurate monitoring of water resource pollution. The experimental results show that the method has high detection accuracy, low leakage rate and high detection efficiency.

Keywords: Convolution neural network; Water resource; Microwave wireless transmission; Pollution detection

1. Introduction

The lack of detection data cannot truly reflect the change of water pollution and cannot meet the needs of environmental management, thus, the research on water pollution detection methods is imperative [1–3].

In order to solve the defects in the detection process, this paper proposes a method of water pollution detection based on convolution neural network, the main contents are as follows:

The water pollution data after treatment is input into convolutional neural network to realize water pollution detection [4–6].

The validity of the method is verified by the accuracy of the detection results, the rate of missing detection of pollution elements and the detection efficiency.

2. Material and methods

The method to detect water resources pollution based on convolution neural network inputs the processed data into the convolutional neural network to realize the water resources pollution detection. The specific steps are described as follows:

2.1. Determine the input and output parameters of convolution neural network algorithm

According to the actual analysis, the main input and output parameters of convolution neural network model are determined at first [7–9]. Secondly, the sample data is processed. Finally, the statistics of daily data are calculated. The amount of data recorded every day is set as N. The average value is set as $\bar{X}$, and the variance is set as S^2:

* Corresponding author.

1944-3994/1944-3986 © 2023 Desalination Publications. All rights reserved.

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^2 \tag{2}$$

In order to quantitatively express the inconsistency of each data, it is necessary to determine the degree of deviation of each data [10]. Let $V(i)$ be the deviation degree of data:

$$V(i) = \frac{\left|x_i - \bar{x}\right|}{S^2} \tag{3}$$

If $V(i)$ is relatively large, it means that the data is too far from the average data, and the data cannot be received. On the contrary, if $V(i)$ is small, it means that the data is closed to the average data, so that the data can be received [11].

When $V(i) \geq 1.1$, the data was "bad data";
When $V(i) \leq 1.1$, the data was "valid data".

Generally, the following methods are used to normalize any valid data.

$$x = \frac{x_{\text{sj}} - x_{\text{zj}}}{x_{\max} - x_{\min}} \tag{4}$$

where x refers to the normalized parameter, which can be used as the input parameter of network. x_{sj} is the actual number. x_{zj} is the intermediate number. $x_{\max}$ represents the maximum value of the effective data collected on site, and $x_{\min}$ represents the minimum value:

$$x_{\text{zj}} = \frac{x_{\max} + x_{\min}}{2} \tag{5}$$

2.2. Initialize the weights of convolutional neural network

A series of initial values w_{ij} ($i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$) are generated by random method. The population consists of n groups of neural network weights is established. The number of neurons in hidden layer is m. The initial weights are:

$$\left\{\left(w_{11}, w_{12}, \ldots, w_{1m}\right), \left(w_{21}, w_{22}, \ldots, w_{2m}\right), \ldots, \left(w_{m1}, w_{m2}, \ldots, w_{mm}\right)\right\} \tag{6}$$

2.3. Fitness calculation

After a set of training set data is calculated by convolution neural network, the error E of output node is calculated:

$$E = \frac{1}{2}\sum_{i=1}^{p}(t_i - y_i)^2 \tag{7}$$

where p is the number of output nodes, y_i is the network calculation result, and t_i is the actual value.

The fitness value of convolution neural network weights can be expressed as follows:

$$f = \frac{1}{E+1} \tag{8}$$

Obviously, $0 < f < 1$. The smaller the error E is, the closer to 1 the f is, so the adaptive value of the improved network structure will be larger. For the members in new group, the fitness value $f(x_{ji}^{(OJ)})$ of member is calculated. Meanwhile, the variable with the largest fitness in the family is selected as the next generation population.

Binary coding is performed on new population, so that selection, crossover and variation can be applied to the weights of the corresponding convolutional neural network [12]. In addition, the convolution neural network is trained at first, and then the binary system is decoded to calculate the weights of decimal neural network.

The following formula is the relationship between population coding and population decoding:

$$\left\{\left(w_{11}, w_{12}, \ldots, w_{1m}\right), \left(w_{21}, w_{22}, \ldots, w_{2m}\right), \ldots, \left(w_{m1}, w_{m2}, \ldots, w_{mm}\right)\right\} \tag{9}$$

2.4. k–j mean clustering

After the number m of neurons in hidden layer of convolution neural network is determined, m samples are selected from the input samples x_j ($j = 1, 2, \ldots, N$) as the clustering center. Meanwhile, the input samples need to be classified by the proximity principle.

The x_j ($j = 1, 2, \ldots, N$) satisfying $d_i = \min_i \left\|x_j - c_i\right\|$ and ($j = 1, 2, \ldots, N$; $i = 1, 2, \ldots, M$) is assigned to the multi-input cluster set θ_i ($i = 1, 2, \ldots, M$) of center c_i ($i = 1, 2, \ldots, M$), where $x_j \in \theta_i$.

Firstly, it is necessary to calculate the cluster center c_i, and then calculate the average value of θ_i:

$$c_i = \frac{1}{k}\sum_{j=1}^{K} x_j, \left(j = 1, 2, \ldots, K\right) \tag{10}$$

2.5. Calculation from input layer to hidden layer

Gaussian function is taken as radial basis function:

$$R_i(x) = \exp\left[-\frac{\left\|x - c_i\right\|^2}{2\sigma_i^2}\right], i = 1, 2, \ldots, m \tag{11}$$

where x is the n-dimensional input vector. m is the number of neurons in hidden layer. c_i is the center of the ith basis function. σ determines the width of the basis function around the center. σ has a certain relation with the number m of neurons in the hidden layer. In order to ensure appropriate shape of Gaussian function, the following formula is used to calculate σ:

$$\sigma = \frac{d_m}{\sqrt{2m}} \tag{12}$$

where d_m is the maximum distance of the selected center.

2.6. Calculation from hidden layer to output layer

The initial weights and the output of hidden layer are used to calculate the linear mapping from hidden layer $R_i(x)$ to output layer y_i. That is:

$$y_i = \sum_{i=1}^{m} w_{ij} R_i(x), k = 1,2,\ldots,p \tag{13}$$

where p is the output node.

2.7. Selection

The corresponding population fitness f_i and the total population fitness F are calculated:

$$F = \sum_{i=1}^{N} f_i \tag{14}$$

The selection probability p_i and cumulative probability q_i of each population were calculated:

$$p_i = \frac{f_i}{F} \tag{15}$$

$$q_i = \sum_{j=1}^{i} p_j \tag{16}$$

After rotating N times, we can select N populations. The steps on computer include: generate random number r between (0,1). If $r < q_i$, the first population v_1 will be selected, otherwise v_2 ($2 \le i \le m$) will be selected to meet $q_{i-1} < r < q_i$. Obviously, the larger the fitness, the greater the probability of being selected.

2.8. Crossing

Reset crossover probability p_c:

$$p_c = \frac{f_{\max} - f'}{f_{\max} - \overline{f}} \tag{17}$$

where f' is the maximum fitness of the crossover population in the parent population. $\overline{f}$ is the average fitness value of populations.

3. Results and discussion

In order to verify the effectiveness of the method to detect water resources pollution based on convolutional neural network, it is necessary to test this method. This test needs to be completed in Simulink platform [13]. The water pollution detection method based on convolutional neural network (method 1), the water resource pollution

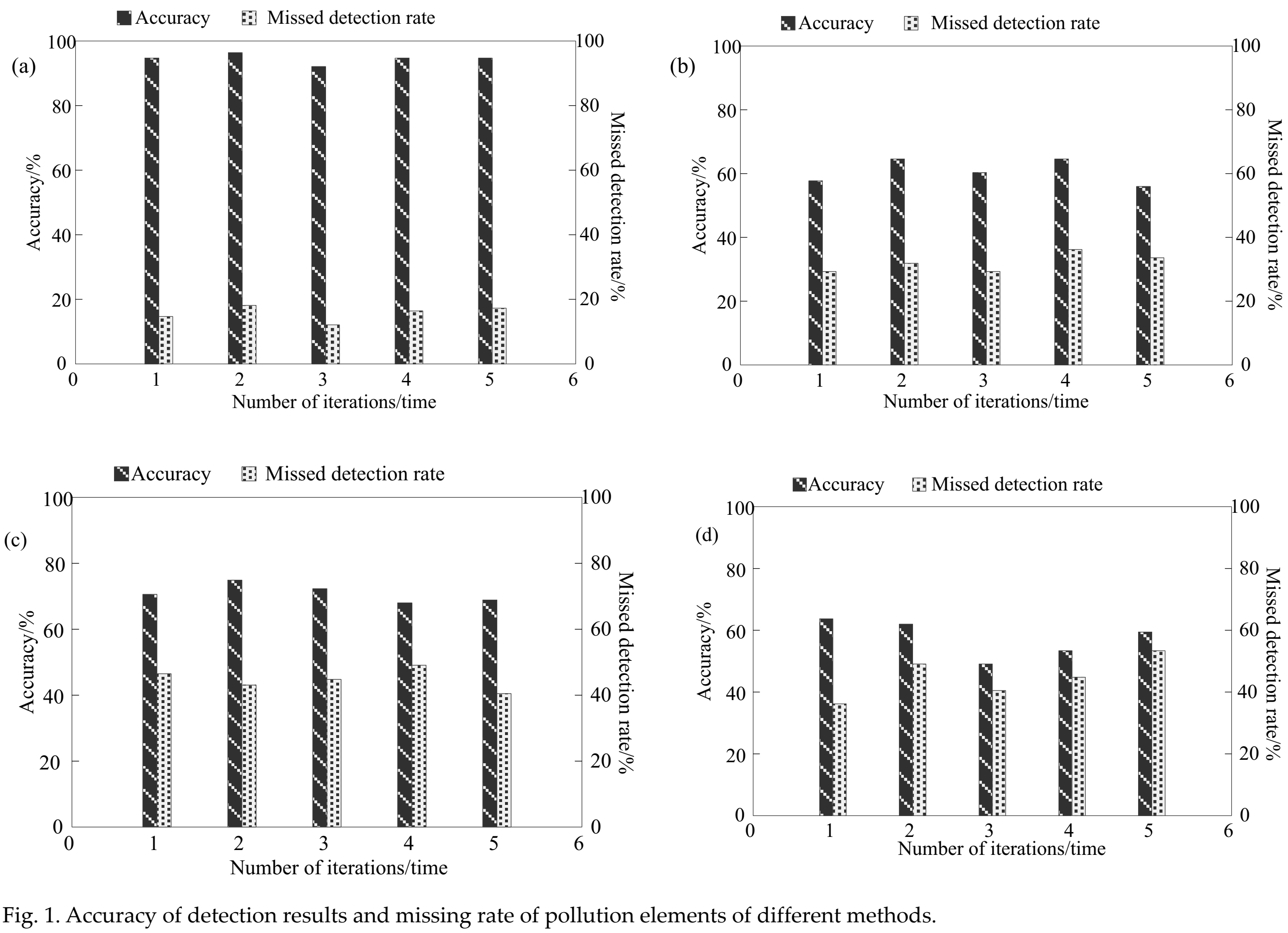

Fig. 1. Accuracy of detection results and missing rate of pollution elements of different methods.

detection method based on supervised learning (method 2), the water resources pollution detection method based on Bayesian formula (method 3) and the water resources pollution detection method based on EEMD (method 4) were adopted. The accuracy of detection results and the missing detection rate of pollution elements of the four methods were compared. The test results are shown in Fig. 1.

By analyzing the data in Fig. 1, we can see that the accuracy of detection results in multiple iterations obtained by method 1 is higher than that of methods 2, 3 and 4 [14,15]. The missed detection rate of pollution elements in multiple iterations obtained by method 1 is lower than that of methods 2, 3 and 4.

4. Conclusion

Due to the problems of low accuracy of detection results, high rate of missing detection of pollution elements and low detection efficiency of existing water resources pollution detection methods, a method of water resources pollution detection based on convolution neural network is proposed. In this method, Kalman filter algorithm is used to process the collected data, and convolution neural network is used to detect water pollution, which can accurately complete the detection of water pollution in a short time.

Acknowledgments

Judicial Administration Research Project of Guangdong Province in 2020: Design and Application of Judicial Big Data Platform System (No. GDSFT20061); Judicial Administration Research Project of Guangdong Province in 2020: Big Data Risk Analysis and Prevention Strategies (No. GDSFT20062); Judicial Administration Research Project of Guangdong Province in 2020: Innovative Research on Accurate Law Popularization Model Driven by Big Data Mining Technology (No. GDSFT20122).

References

[1] X.G. Zhao, B. Li, Detection and evaluation of surface water pollution in tailing reservoir area of Chen'er gold mine, Coal Technol., 36 (2017) 187–188.
[2] Q.Y. Zhao, L.J. Li, H.W. Jin, L. Jiang, Z.X. Du, Water pollution and determination methods of antibiotics in aquaculture, Chin. J. Anal. Lab., 37 (2018) 103–107.
[3] F.H. Pan, Y.B. Wang, X.X. Zhang, J.S. Huang, Y.B. Jiang, X.B. Hu, S. Li, Comparison of pollution diffusion between river mouth inflow (RMI) and beach uniform inflow (BUI) in sudden water accident, Water Environ. J., 33 (2019) 111–123.
[4] H.H. Li, T. Zhang, D.C.W. Tsang, G.X. Li, Effects of external additives: biochar, bentonite, phosphate, on co-composting for swine manure and corn straw, Chemosphere (Oxford), 248 (2020) 125927, doi: 10.1016/j.chemosphere.2020.125927.
[5] S. Wang, K. Zhang, L.P.H. van Beek, X. Tian, T.A. Bogaard, Physically-based landslide prediction over a large region: scaling low-resolution hydrological model results for high-resolution slope stability assessment, Environ. Modell. Software, 124 (2020) 104607, doi: 10.1016/j.envsoft.2019.104607.
[6] T. Zhang, X.Y. He, Y.X. Deng, D.C.W. Tsang, H.M. Yuan, J.B. Shen, S.C. Zhang, Swine manure valorization for phosphorus and nitrogen recovery by catalytic–thermal hydrolysis and struvite crystallization, Sci. Total Environ., 729 (2020) 138999, doi: 10.1016/j.scitotenv.2020.138999.
[7] B. Chen, Y.N. Wang, Study on establishment of collaboration mechanism of water resources conservation and water pollution control, Yangtze River, 47 (2016) 10–13.
[8] S.Y. Tao, G.L. Wang, Distribution characteristics and health risk assessment of main pollutants in different types of water sources, Water Resour. Power, 34 (2016) 10–14.
[9] Z.-h. Yin, Y. Zhu, J. Zhang, X.-l. Zhang, J. Zhang, Enhanced Raman experiments of graphene-Ag nanoparticles prepared with annealing method, Spectrosc. Spectral Anal., 39 (2019) 165–173.
[10] S.-s. Zhang, H.-h. Liu, J. Qiang, X.-k. Liu, X.-q. Zhu, Synchronous inversion of groundwater pollution source and aquifer parameters based on Bayesian formula, China Environ. Sci., 39 (2019) 226–236.
[11] H.Q. Fang, H.F. Xue, Y.Z. Jiang, T. Zhou, Y. Wan, H. Wang, Outlier detection and correction for water resources monitoring data based on EEMD, Trans. Chin. Soc. Agric. Mach., 48 (2017) 257–263.
[12] J.W. Liu, J.C. Zhang, W.L. Ma, B. Liu, Y.Z. Zhao, Pollution and distribution characteristics of bioaerosol in municipal wastewater treatment plant, Ecol. Environ. ENCEs, 22 (2013) 657–661.
[13] L. He, J. Shen, Y. Zhang, Ecological vulnerability assessment for ecological conservation and environmental management, J. Environ. Manage., 206 (2018) 1115–1125.
[14] K. Zhang, Q.Q. Wang, L.J. Chao, J.Y. Ye, Z.J. Li, Z.B. Yu, T. Yang, Q. Ju, Ground observation-based analysis of soil moisture spatiotemporal variability across a humid to semi-humid transitional zone in China, J. Hydrol. (Amsterdam), 574 (2019) 903–914.
[15] W.J. Yang, Y. Zhao, D. Wang, H.H. Wu, A.J. Lin, L. He, Using principal components analysis and IDW interpolation to determine spatial and temporal changes of surface water quality of Xin'anjiang river in Huangshan, China, Int. J. Environ. Res. Public Health, 17 (2020) 2942, doi: 10.3390/ijerph17082942.