Hindawi
Scientific Programming
Volume 2021, Article ID 8271950, 11 pages
https://doi.org/10.1155/2021/8271950

Research Article

Water Pollution Prediction Based on Deep Belief Network in Big Data of Water Environment Monitoring

Li Liang

Chongqing Industry Polytechnic College, Chongqing 401120, China

Correspondence should be addressed to Li Liang; swaumyh@swu.edu.cn

Received 11 November 2021; Revised 30 November 2021; Accepted 2 December 2021; Published 22 December 2021

Academic Editor: Ahmed Farouk

Aiming at the problems that the traditional water quality prediction model is generally not high in prediction accuracy and robustness, a water pollution prediction using deep learning in water environment monitoring big data is proposed. *Objective*. To optimize and improve the prediction accuracy of the water quality prediction model. Firstly, in the water environment monitoring system, the Internet of Things big data technology is used to accurately sense and monitor the real-time data of sewage treatment equipment and sewage quality. Then, the deep belief network (DBN) is used to build the water pollution prediction model, and the collected sewage treatment data is analyzed to predict the water quality status. Finally, particle swarm optimization algorithm is used to dynamically optimize the number of hidden layer neural units and learning rate in the DBN prediction model, which makes the prediction results more scientific and accurate. Based on the sampling data of Shanghai Jinze Reservoir, the proposed model is experimentally analyzed. The results show that the probability of accurate location of the pollution source is not less than 70%. And under the two indicators of chemical oxygen demand and biological oxygen demand, the root mean square error and correlation coefficient are 3.073, 0.9892 and 1.958, 0.9565, respectively, which are better than other comparison models.

1. Introduction

In recent years, with the rapid development of cities and social economy, the issue of water resources has gradually become a hot social issue. In association with this, the problem of water pollution is particularly prominent, which is directly related to the long-term development of my country's economy and society [1]. As shown in Figure 1, water eutrophication caused by industrial waste water, domestic sewage, accidental pollution source leakage, and other reasons, such as serious excess of toxic and hazardous substances, and other water resources problems are common [2].

However, in real life, the detection of water resources is still done manually in a considerable part of the area and submitted to the laboratory for analysis. Although it is possible to obtain as detailed water quality information as possible, it will greatly consume manpower and material resources, and it is difficult to ensure timeliness [3]. At present, sensors are widely used to collect water quality data of the water supply pipe network in real time and transmit the data to the server through the network for centralized analysis. The method of using sensor data to detect and trace the source of water pollution is directly related to the choice of sensor type, usually including sensors for specific pollutants and general-purpose sensors. As far as a specific sensor is concerned, it has better performance in detecting specific pollutants, but its ability to detect other pollutants is weak. Usually this type of sensor is mainly aimed at pollutants such as heavy metal ions [4]. For general-purpose sensors, they are not designed for a specific type of pollution, so they have a more general detection ability for most pollution types [5]. Faced with the massive detection data generated by many sensors in the water supply network, its analysis and judgment also require updated technical support. Through the study of water resources prediction models, the use of water environment monitoring big data to predict the pollution of water sources is the key research direction [6, 7].

At present, scholars at home and abroad have done a lot of research on water quality prediction based on sensor big data. Traditional water quality prediction models mainly

Figure 1: Typical water pollution cases.

include time series models, regression analysis models, and grey system theory models [8, 9]. Reference [10] proposed a recurrent neural network water quality prediction method based on sequence-to-sequence framework. The gate loop unit model is used as the encoder and decoder, and the factorization machine is integrated in the model to solve the problem of high sparsity and high-dimensional feature interaction in the data. However, it cannot accurately predict data with large fluctuations. Reference [11] proposed a method to estimate the concentration of environmental pollutants in water based on environmental parameters. Symbolic constraints are used to express domain knowledge, and the influence of symbolic constraints on prediction performance is studied by using censored data sets. Its prediction accuracy is greatly affected by the data itself, and it is only suitable for medium- and short-term prediction. Reference [12] proposed a prediction model based on nonlinear regression for the problem of irrigation water quality. It has flexible and accurate evaluation performance for irrigation water quality. Traditional forecasting models often only pay attention to the characteristics of the data itself, without fully considering the interrelationship between the data. The prediction accuracy is generally not high, and it is difficult to accurately predict and monitor the water quality parameters of the water environment [13].

With the continuous improvement of the computing performance of smart hardware, deep learning and artificial intelligence have developed rapidly. They are continuously integrated into all aspects of national life and industrial control [14]. As an important component in the field of deep learning, cyclic neural network fully considers the long-term dependence of time series data and can handle time series data well [15]. Reference [16] proposed a water quality parameter analysis and water quality prediction method using linear regression analysis and artificial neural network. The artificial neural network has a good forecasting effect, but linear regression analysis cannot be used for nonlinear forecasting. Reference [17] proposed a seawater quality prediction method based on artificial neural network and multiple linear regression model. The seawater quality of mangroves and estuaries has been accurately predicted. However, a large amount of sample data is required for training, and the parameters set by experience can easily lead to the appearance of local extreme values. Reference [18] combines convolutional neural network and long-short-term memory model to predict water quality, which has good accuracy and predictive performance. However, the training sample should not be too large, and it is more sensitive to missing data. Reference [19] proposed a water supply and drainage health monitoring method combining fog computing and cloud computing based on the Internet of Things water supply system, which improved the prediction accuracy. However, the mining of the in-depth correlation information between data is not deep enough, and the utilization rate of monitoring big data needs to be further improved.

Aiming at the problem that traditional prediction models cannot handle massive data from multiple sensors, a water pollution prediction model using deep learning in the big data of water environment monitoring is proposed. The innovations of the proposed model are summarized as follows:

(1) Due to the lack of data in most waters and the unclear water management mechanism, the accuracy of traditional prediction models is not high. The proposed model introduces deep learning technology, which has good data nonlinear approximation, self-learning, and generalization capabilities and can achieve a more ideal water quality prediction effect.

(2) The particle swarm optimization algorithm is used to dynamically optimize the number of hidden layer neural units and the learning rate in the prediction model. In order to improve its convergence speed and generalization ability, the prediction results are more scientific and accurate.

2. Related Technology

2.1. Deep Belief Network. Deep belief network (DBN) is a directed graph model widely used at present. It can be seen as a superposition of multiple restricted Boltzmann machines (RBM). First, the effective unsupervised greedy layer-by-layer training method is used to initialize the DBN weights; that is, only two adjacent layers are trained each time, and each output is used as the input of the next training, and the training is performed layer by layer. The features are extracted from the input sample data to obtain the parameters of the global network model. Then the supervised learning method is used to fine-tune all the parameters, further optimize the network, and get the trained DBN. The DBN network structure is shown in Figure 2.

In order to reduce the complexity of the algorithm, the whole DBN is divided into several RBMs. The RBM was trained layer by layer with the fast training method of contrast divergence (CD). Although CD algorithm does not follow any function gradient and its maximum likelihood estimation is not accurate, it is very effective for training depth structure similar to DBN [20, 21]. The RBM training process optimizes the initial parameters of the network model to avoid the situation where the model falls into a local extreme value due to improper initial values. Finally, the back propagation (BP) algorithm is used to supervise and fine-tune the network parameters. This is a local space search, so the speed is faster, and it is not easy to fall into a local extreme value situation.

2.2. Particle Swarm Optimization Algorithm. Particle swarm optimization (PSO) algorithm is a swarm intelligent optimization algorithm that simulates the collective cooperation of birds to find food. It was first proposed by J. Kennedy and R. Eberhart in 1995. PSO combines the advantages of swarm intelligence optimization algorithms and the advantages of evolutionary calculations to achieve global optimal search in complex spaces.

In the PSO algorithm, in order to achieve the optimality of the behavior of the entire group, the individual is represented by particles that specify the corresponding behavior rules. The particles find the optimal position based on their own experience and group experience and constantly update themselves, and the particles find the optimal solution through cooperation and mutual assistance.

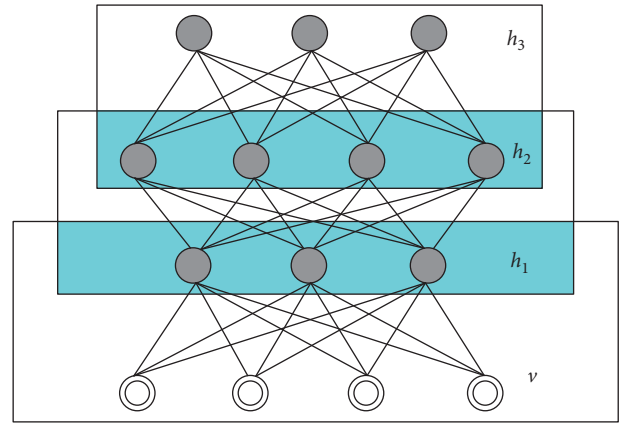


Figure 2: The structure of deep belief network.

Mathematical expression of PSO: in the D-dimensional search space, n represents the number of particles $i = 1, 2, \ldots, n$. The position of the i-th particle is denoted by $X_i = (x_{i1}; x_{i2}; \ldots; x_{i\,D})$. The historical optimal position of the i-th particle is represented by $P_i = (p_{i1}; p_{i2}; \ldots; p_{i\,D})$. The velocity of the particle is denoted by $V_i = (v_{i1}; v_{i2}; \ldots; v_{i\,D})$. The particle velocity and position update formula are as follows:

$$\begin{cases} V^{k+1} = \omega V^k + c_1\gamma_1\left(P_i^k - X^k\right) + c_2\gamma_2\left(P_g^k - X^k\right) \\ X^{k+1} = X^k + V^{k+1} \end{cases}, \quad (1)$$

where the right side of the speed update formula is inertial part, cognitive part, and social part. ω is the inertia weighting factor, which is generally between (0.2, 0.9). c_1 and c_2 are learning factors, generally take the same normal number between (0,4), and usually take 2. γ_1 and γ_2 are random positive numbers, evenly distributed between (0,1). P_g represents the historical global optimal solution. Sometimes in order to limit the speed of the particles, the upper limit $V_{\max}$ and the lower limit $V_{\min}$ of the particle speed are set according to different situations, generally set to 2.048 and -2.048.

3. System Structure

In the water quality pollution prediction system architecture based on deep learning in the water environment monitoring big data, the overall topological structure and functional structure of the system are mainly designed, and the overall design and implementation of the system are planned [22]. The system design goals mainly include two aspects: water quality data collection based on big data of the Internet of Things, water quality pollution prediction and control based on deep learning. Its overall topological structure is shown in Figure 3.

The system uses wireless sensor nodes as data sensing equipment for big data of the Internet of Things to monitor the sewage water quality of sewage treatment equipment and every intermediate link in the sewage treatment process. The monitoring results are handed over to the cloud computing storage platform. On the cloud computing storage platform,

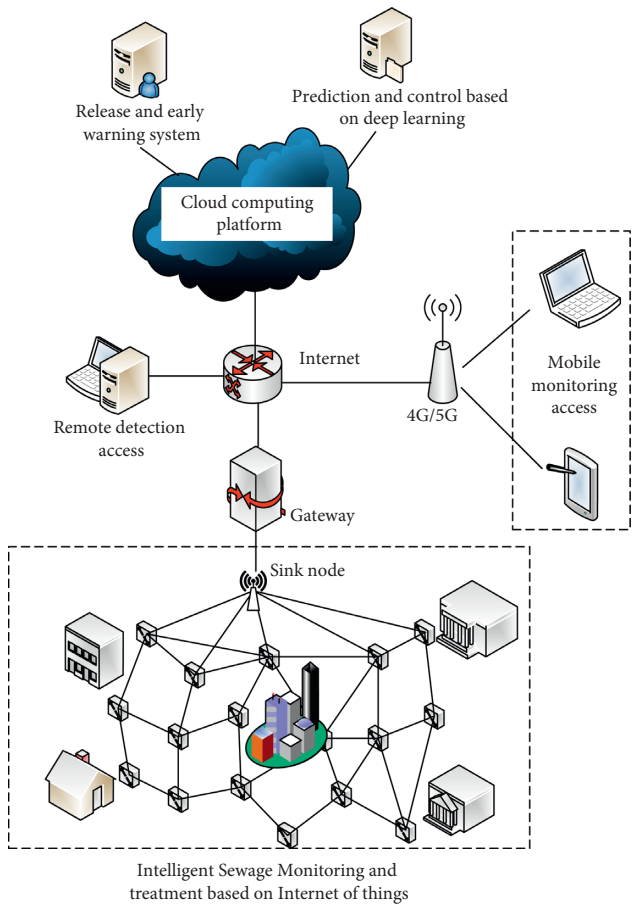


Figure 3: Overall topological structure of the system.

deep learning algorithms are used to process and analyze monitoring data, generate prediction results, and intelligently control sewage treatment equipment.

General sewage treatment is divided into four levels. The sewage treatment perception system based on big data of Internet of Things deploys wireless sensor nodes for each treatment process, each important parameter index, and important treatment equipment of sewage treatment, realizes the information perception of the whole sewage treatment process without dead angle and full coverage, and ensures the comprehensive, in-depth, and thorough perception and collection of the sewage treatment process information. Each sensor node self-organizes to build a wireless multihop network through the Zigbee protocol and reports the collected sewage treatment data to the sink node. The coordinator node will upload the data to the cloud platform through the gateway for subsequent intelligent analysis and processing.

4. Intelligent Prediction and Control Design of Sewage Treatment Process Based on Deep Learning

4.1. Algorithm Design of Prediction Model. The intelligent prediction and control of sewage treatment based on deep learning are as follows. Firstly, the water environment monitoring IOT collects the parameters of each treatment equipment and the intermediate sewage quality data in the sewage treatment process and obtains the historical original data. After data preprocessing, the training data set is obtained. Then, the unsupervised learning machine learning DBN is used to model the wastewater treatment prediction. The optimal network structure of DBN is determined by using the training data set, including the number of nodes in the input layer, the number of nodes in the hidden layer, and the number of layers in the hidden layer, and the weight is adjusted. The training data is used to train this model, and the final model is adjusted continuously.

The big data system of water environment monitoring collects the data of sewage treatment equipment parameters and intermediate sewage quality data in real time to form the current original data. At present, the original data becomes the prediction data set after data preprocessing. The DBN model and prediction data set are used to predict the results of sewage treatment and then control the operation of related equipment in the whole process of sewage treatment.

DBN overlaps multiple RBM models together, regards the visual layer of each RBM model as the input layer and the hidden layer as the output layer, and then completes the training [23]. The visual layer of the network and the hidden layer unit are interconnected (no connection within the layer), and the hidden layer unit can obtain the high-order correlation of the input visual unit. Compared with the traditional Sigmod reliability network, the learning of RBM weights is relatively easy [24]. In order to obtain generative weights, unsupervised greedy layer-by-layer implementation is used in pretraining [25, 26]. In the training process, the Gibbs sampling principle is adopted; that is, the visible vector value is mapped to the hidden layer unit. Then the visible unit is reconstructed from the hidden layer unit. These new visual units are mapped to hidden layer units again, and new hidden layer units are obtained. A typical DBN network with only one hidden layer can use the joint probability density distribution to describe the relationship between the input vector x and the hidden vector g^i. The mathematical expression is as follows:

$$\begin{aligned}\rho\left(x, h^1, h^2, \ldots, h\right) &= \rho\left(x \mid h^1\right) P\left(h^1 \mid h^2\right), \\ &\rho\left(h^2 \mid h^3\right) \ldots \rho\left(h^i \mid h^{i+1}\right) \rho\left(h \mid h\right),\end{aligned} \tag{2}$$

where $\rho(h^i|h^{i+1})$ is the conditional probability distribution. Think of the hidden layer h^i as a random binary vector with n^i elements h^i_j:

$$\begin{aligned}\rho\left(h^i \mid h^{i+1}\right) &= \prod_{j=1}^{n^i} \rho\left(h^i_j \mid h^{i+1}\right) \rho\left(h^i_j = 1 \mid h^{i+1}\right) \\ &= \operatorname{sigm}\left(b^i_j + \sum_{k=1}^{n^{i+1}} \omega^i_{kj} h^{i+1}_k\right),\end{aligned} \tag{3}$$

where $\operatorname{sigm}(t) = 1/(1+e^{-t})$, b^i_j is the bias value of the j th unit in the i-th layer, and ω^i is the weight of the i-th layer.

After training, you need to fine-tune the DBN training. According to the loss function of the input data and the reconstructed data, the BP algorithm is used to fine-tune the

correlation network parameters to minimize the loss function. The formula of the loss function is

$$L(x - x') = \|x - x'\|_2^2. \quad (4)$$

Among them, x is the true value of the training data, and $x\prime$ is the value of the DBN fitting function.

Considering that the final discharged water quality is the final direct indicator, the final water quality is used as the indicator of the sewage treatment result. Realize the prediction of the discharge water quality by constructing the DBN network. The construction process of the sewage treatment forecast DBN network model is shown in Figure 4.

The input of DBN water pollution prediction model includes the parameters of each process and the sewage quality after each process. And through the bottom-up combination of multiple RBMs to build a DBN network, the final output of sewage quality is obtained.

4.2. Optimization of Model Parameters for Water Quality Prediction. When designing the structure of the DBN, it is necessary to determine the number of hidden layers, the number of nodes contained in each hidden layer, the learning rate of RBM, and the learning rate of the fine-tuning process. The setting of these parameters largely affects the prediction performance of the water pollution prediction model. However, there is no relevant theory to clarify the best selection method for these parameters. Many researchers use a large number of experimental comparisons, experiences, and trial-and-error methods to determine these parameters and choose a better network structure of the water pollution prediction model. The network structure needs to be readjusted every time the impact factor related to the forecast changes. The network structure of each model is only suitable for a specific environment, resulting in poor generalization ability of the predictive model. Moreover, the accuracy of the model's prediction results is related to the model user, and experienced experts may get better results.

The proposed model uses PSO to optimize the parameters of the water pollution prediction model. PSO can not only avoid falling into local extreme values but also ensure the global search ability and can optimize the parameters of the water pollution prediction model. Regarding each parameter to be optimized as a particle, it iteratively adjusts continuously to continuously approach the global optimal solution. The convergence speed is fast, and the adaptability of the water pollution prediction model becomes stronger, and the generalization ability is improved. The process of using the PSO algorithm to dynamically optimize the water pollution prediction model based on deep learning is shown in Figure 5.

The number of hidden layer units of the DBN network is M, and the learning rates of RBM1, RBM2, and RBM3 are ε_1, ε_2, and ε_3, respectively. The particle $x(m, \varepsilon_1, \varepsilon_2, \varepsilon_3)$ is a four-dimensional vector, where $\varepsilon_1, \varepsilon_2, \varepsilon_3 \in (0, 1)$.

The pseudocode of the specific PSO algorithm is shown in Algorithm 1.

5. Experiment and Analysis

The water quality data in the experiment comes from the real monitoring data of the main water quality indicator chemical oxygen demand (COD) of Shanghai Jinze Reservoir, the main water source in Shanghai, from April 30, 2019, to November 30, 2019, and according to the collection frequency per minute to obtain 300,520 monitoring value data. The original COD data measurement value is shown in Figure 6.

5.1. Data Preprocessing

5.1.1. Missing Value Processing. In the original data set, some sensor data is missing at some moments. Therefore, it is necessary to fill in the missing data before using the data. Since the sensor data has a high probability to remain relatively stable in a short period of time, the forward filling method is adopted. That is to fill in the missing data at the current moment based on the data at the previous moment. The possible data missing in the sensor data collected at 4 consecutive times within a certain period of time and the corresponding forward filling results are shown in Table 1. The black data is the real measurement data, and the red data is the filling data.

5.1.2. Standardization. For the learning of multidimensional feature data, standardization can often facilitate data processing and speed up the convergence speed of model training. For the data set used, each piece of data has different characteristics. Therefore, it is also necessary to standardize the original data set, so as to avoid the phenomenon that the model convergence is too slow due to the large difference between the original data of different characteristics. The standardization method adopted is that, for each sensor indicator, if the mean value of the indicator in the set time window is recorded as μ, the standard deviation is recorded as δ. Then the observed value x_t of this indicator at time t becomes

$$\hat{x}_t = \frac{x_t - \mu}{\delta}. \quad (5)$$

After this transformation, the values between different sensors are scaled to a range that can be directly compared. Moreover, using the transformed data to train the model can also speed up the convergence speed of the model training to a certain extent.

5.2. Evaluation Index. Two indicators, RMSE and correlation coefficient R, are used to evaluate the performance of the water quality prediction model.

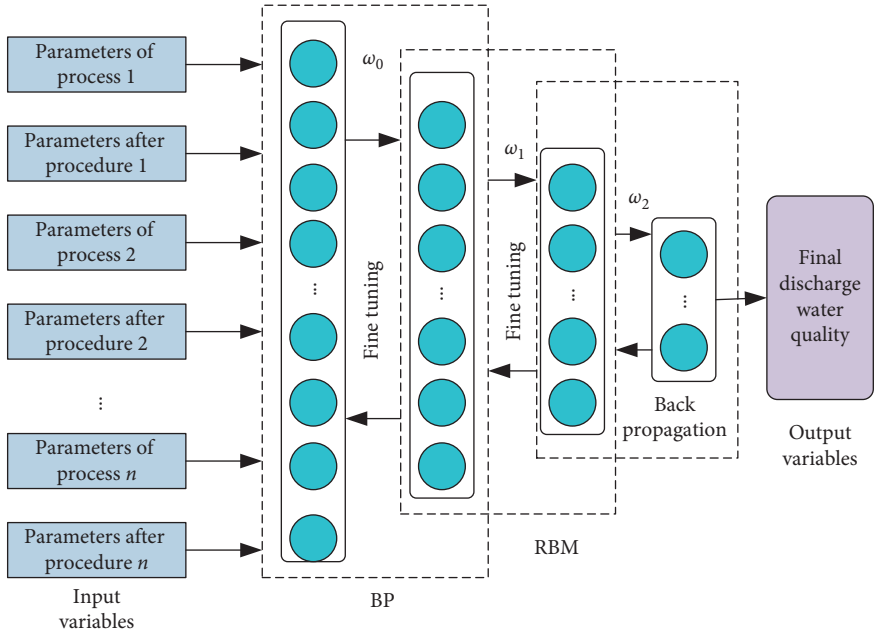


Figure 4: Training process of DBN network model for wastewater treatment prediction.

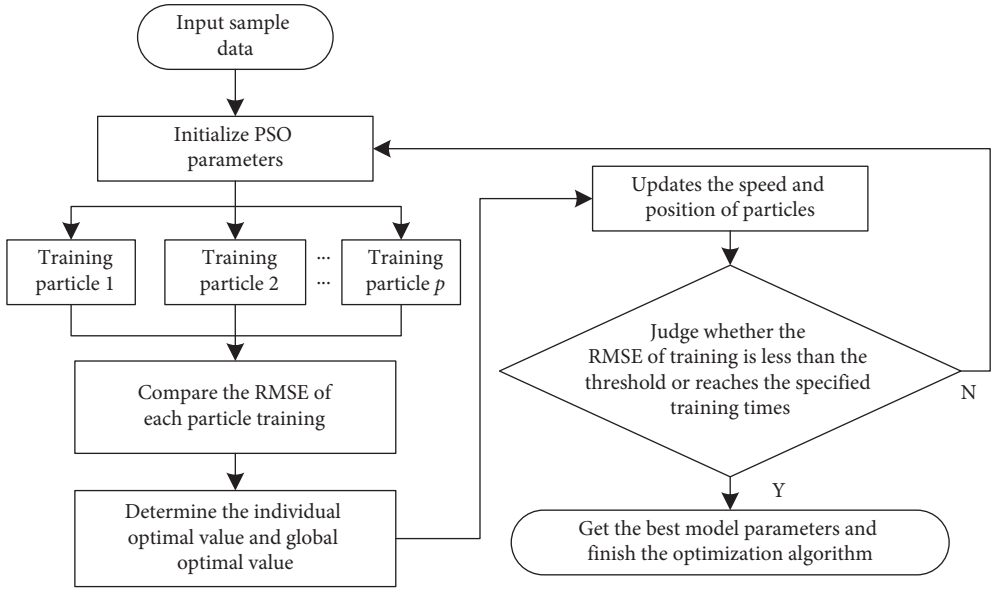


Figure 5: Optimization process of PSO algorithm.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - y_i'\right)^2}{n}},$$
$$R = \frac{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)\left(y_i' - \overline{y}\prime\right)}{\sqrt{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2 \cdot \sum_{i=1}^{n}\left(y_i' - \overline{y}'\right)^2}}. \tag{6}$$

In the formula, y_i represents the actual data of the i-th sample, and y_i' represents the predicted data of the i-th sample. $\overline{y}$ and $\overline{y}\prime$, respectively, represent the average value of n actual data and the average value of n predicted data. When using RMSE and R as indicators to evaluate the water quality prediction model, the smaller the RMSE, the better, and the closer R to 1, the better.

5.3. Pollution Source Location Based on Deep Learning Model. First, the data set obtained by monitoring is divided into two data sets to evaluate the effect of the deep learning model in the first stage of traceability in the location of pollution sources. Before that, the same segmentation is performed on the two data sets: 80% of the data is used as the training set, and 20% of the data is used as the test set. The positioning results of the

(1) Parameter: N:The population size of the particle swarm; C:The maximum number of iterations.
(2) Begin
(3) Initialize the particle's position $x^{c=0}$ and velocity $v^{c=0}$.
(4) The root mean square error (RMSE) between the predicted value and the expected value is used to find the optimal position $x^c_{i_{\text{pbest}}}$ and the global optimal position x^c_{gbest} of each particle.
(5) Update speed and position information. Calculate and update the speed and position information of the particles according to the speed update formula and the position update formula: $\begin{cases} x_i^{c+1} = x_i^c + v_i^{c+1} \\ v_i^{c+1} = \omega v_i^c + c_1\gamma_1\left(x^c_{i_{\text{pbest}}} - x_i^c\right) + c_2\gamma_2\left(x^c_{\text{gbest}} - x_i^c\right) \end{cases}$
(6) If RMSE convergence and $c = C$ then PSO algorithm ends;
(7) otherwise $c = c + 1$, return to step 3.
(8) End

Algorithm 1: Pseudocode of PSO algorithm.

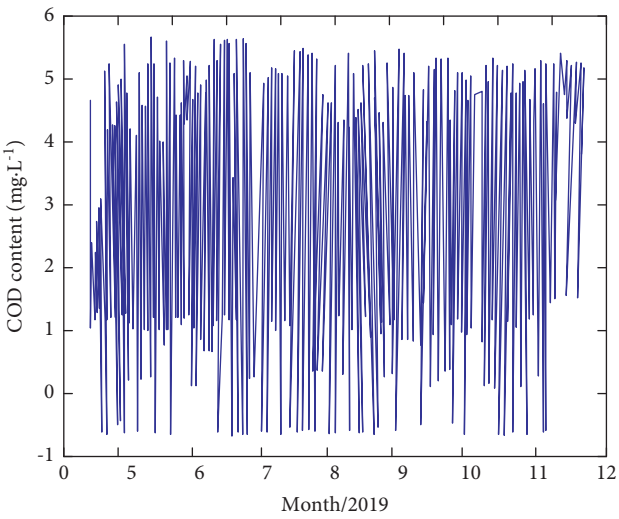


Figure 6: Measured value of original COD data.

Table 1: Forward fill results with missing data.

Time	Cl	Cl_2	pH	Tp	COD
...	...	...	...	...	...
T	A_t	B_t	C_t	D_t	E_t
$t+1$	A_{t+1}	B_t	C_{t+1}	D_t	E_{t+1}
$t+2$	A_{t+1}	B_{t+2}	C_{t+2}	D_t	E_{t+2}

proposed model on the two data sets are shown in Tables 2 and 3. It should be noted that T in the table represents the number of samples. That is, the data used in training and testing the model is the concentration data at T moments after the pollution occurs, rather than all the data during the entire simulation period. For example, when the number of sensors is 4 and that of T is 12, it means that the data used by the corresponding proposed model is the data collected at the 4 sensor nodes at 12 moments after the pollution occurred.

Since the proposed model will output multiple suspicious pollution source nodes at the same time, the evaluation of the model can be described by whether the actual pollution source node is included in several nodes predicted by the proposed model. Therefore, it can be seen from Tables 2 and 3 that, (1) under the combination of the number of various sensors and the number of samples, the 6 nodes predicted by the proposed model obtained on the two data sets have a high probability of containing the real pollution source nodes. In the worst case, the probability also exceeds 70%. Therefore, the constructed DBN model for locating pollution sources is effective. (2) The effect of the proposed model is still reliable even when the number of samples T is small. Therefore, the method of locating pollution sources based on the DBN model does not rely on long-term data collection. The process of locating the pollution source is a task that can be completed in a relatively short period of time.

Table 2: The probability that the 6 nodes predicted by the proposed model in dataset A contain real pollution source nodes.

Number of sensors	$T=2$	$T=6$	$T=8$	$T=12$	$T=14$
4	0.854	0.868	0.892	0.886	0.884
6	0.795	0.807	0.856	0.875	0.863
8	0.733	0.774	0.827	0.861	0.859

Table 3: The probability that the 6 nodes predicted by the proposed model in dataset B contain real pollution source nodes.

Number of sensors	$T=6$	$T=12$	$T=18$
10	0.713	0.740	0.769
20	0.752	0.785	0.791
30	0.868	0.863	0.884

5.4. Comparison of Predicted Value and Actual Value. The water quality pollution prediction model based on deep learning is used to predict this sewage treatment plant dataset, and the result is shown in Figure 7. Among them, the predicted output value of the model includes biological oxygen demand (BOD) in addition to COD.

It can be seen from Figure 7 that whether it is COD or BOD, the predicted value can better approximate the actual value, and the two curves have a higher consistency. The predicted value of individual points differs greatly from the actual value, which may be caused by test errors or other uncertain parameters. Taken together, it can be demonstrated that the proposed model has better predictive performance.

5.5. Comparison with Other Models. In order to better demonstrate the performance of the proposed model, the prediction experiment was performed on the same set of experimental data as the model in [17]. The results of the comparative experiment are shown in Figure 8.

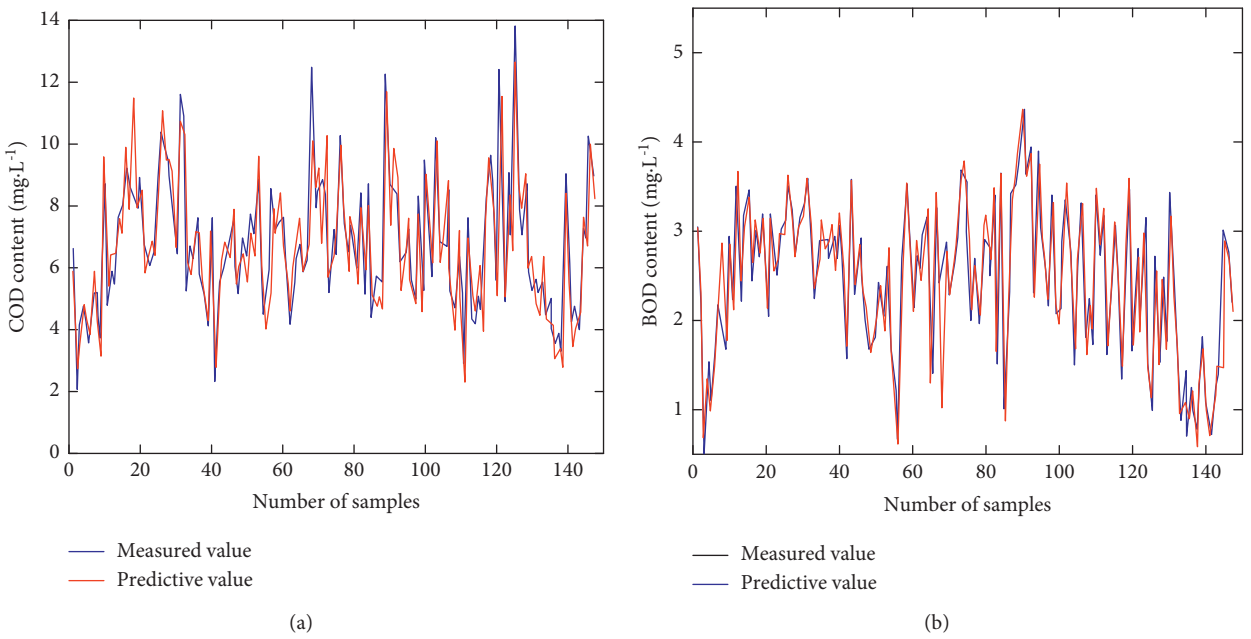


Figure 7: The prediction and output value of different water quality parameters by the model. (a) The predicted and actual values of COD. (b) The predicted and actual values of BOD.

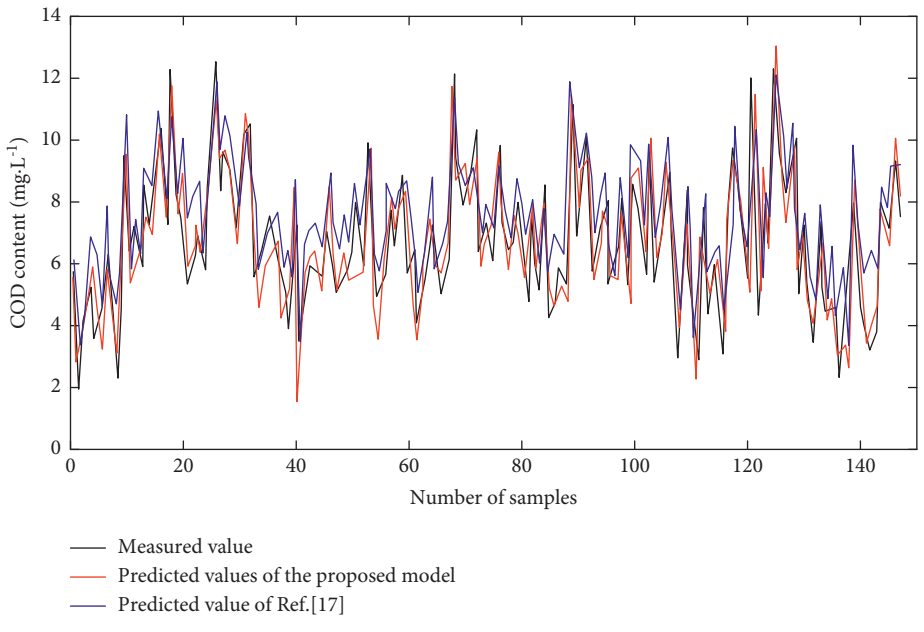
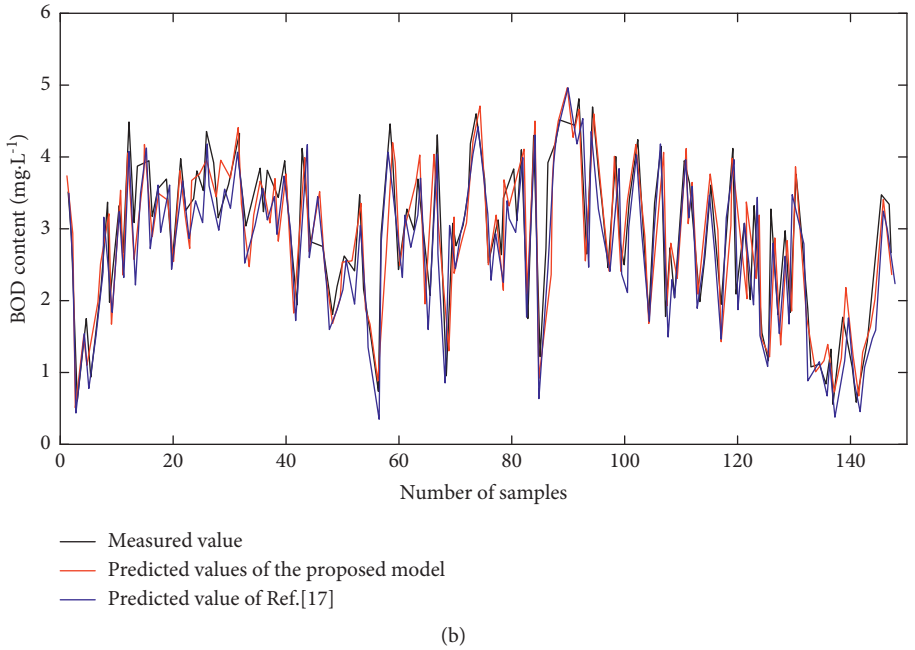


Figure 8: Continued.



(b)

Figure 8: Comparison of predicted value and output value under different models. (a) Comparison of COD predicted value and actual value in different models. (b) Comparison of BOD predicted value and actual value in different models.

Table 4: Comparison results of evaluation indexes of two prediction models.

Model	Water quality index	RMSE	R
Ref. [11]	COD	5.384	0.9533
	BOD	3.190	0.9017
Ref. [17]	COD	4.521	0.9641
	BOD	2.185	0.9239
The proposed model	COD	3.073	0.9892
	BOD	1.958	0.9565

It can be seen from Figure 8 that the water quality prediction model based on the improved DBN and [17] based on the artificial neural network and the multiple linear regression model have good COD and BOD prediction effects, and the curve fits well. Their predicted and actual values are relatively consistent. However, the prediction results of the improved DBN prediction model are closer to the actual data, with smaller errors and higher accuracy.

In order to quantitatively compare and analyze the performance of the two models, RMSE and correlation R are used for evaluation. The results are shown in Table 4. In order to increase the persuasive power, the traditional prediction model proposed in [11] is added.

It can be seen from Table 4 that, for the prediction of COD and BOD, the RMSE and R of the proposed prediction model are 3.073 and 0.9892, 1.958 and 0.9565, respectively, which are better than other comparison models. Because it uses PSO to improve DBN, it can quickly obtain high-precision prediction results. The model in [11] is more traditional, and the prediction results are not ideal. Taking the correlation R of BOD as an example, it is only 0.9017. Reference [17] combines artificial neural network and multiple linear regression model to achieve water quality prediction, and their prediction performance has been improved to a certain extent. But the learning performance advantage of artificial neural network is not obvious, because the overall effect is lacking. The particle swarm optimization algorithm is used to dynamically optimize the number of hidden layer neural units and the learning rate in the prediction model. In order to improve its convergence speed and generalization ability, the prediction results are more scientific and accurate. Therefore, the proposed water quality prediction model based on deep learning has a relatively ideal water quality prediction effect and has certain practical application advantages.

6. Conclusion

Water quality is closely related to people's daily life. In order to improve the quality of life, a highly reliable water quality prediction model is necessary. For this reason, a water pollution prediction model using deep learning in water environment monitoring big data is proposed. In the water environment monitoring system, the Internet of Things big data technology is used to accurately sense and monitor the real-time data of sewage treatment equipment and sewage water quality. And the DBN is improved by PSO to build the water pollution prediction model, so as to get the ideal water

quality prediction results. Based on the sampling data of Jinze reservoir in Shanghai, the proposed model is demonstrated experimentally. The results show that the probability of accurate pollution source location is not less than 70%, and the pollution source location can be completed in a short time. In addition, under the two indicators of COD and BOD, the RMSE of the proposed model is 3.073 and 1.958, respectively, which are better than other comparative models. And the correlation coefficient R is 0.9892 and 0.9565, which are very close to 1. The proposed model only uses a specific data set for experimentation. In reality, the benchmarks of water quality in different water supply pipe networks are not the same, and the types of sensors deployed are also different. Therefore, how to obtain data from a real water supply network and design a learning model for pollution detection still needs to be further explored.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this study.

Acknowledgments

This work was supported by Effects of plants on soil mercury migration and its ecological effects in the water-level fluctuation zone of the Three Gorges Reservoir project of science and technology research program of Chongqing Education Commission of China (no. KJQN201803203).

References

[1] H. Yang, B. Y. Liu, and J. H. Huang, "Forecast model parameters calibration method for sudden water pollution accidents based on improved Bayesian-Markov chain Monte Carlo," *Kongzhi yu Juece/Control and Decision*, vol. 33, no. 4, pp. 679–686, 2018.

[2] B. Liu, H. Wang, X. Lei, Z. Liu, and J. Quan, "Emergency operation rules for water-supply reservoirs under uncertainty and risk in dry seasons," *Water Science and Technology*, vol. 18, no. 5-6, pp. 1682–1695, 2018.

[3] N. Y. Hrybova, O. I. Khyzhan, V. I. Maksin, L. O. Kavshun, and O. L. Tankha, "Determination of xenobiotic imidacloprid content in surface waters," *Journal of Water Chemistry and Technology*, vol. 41, no. 5, pp. 313–317, 2019.

[4] M. A. Zazouli, L. R. Kalankesh, and M. Hasanpour, "Forecast optimal fluoride concentration data in drinking water according to the ambient temperature in Golestan, Iran," *Environmental Quality Management*, vol. 28, no. 3, pp. 1–5, 2019.

[5] B. Kurbanov, A. Primov, and B. Kurbanov, "Analysis and forecast of the ecological conditions of surface water in Uzbekistan," *InterCarto. InterGIS*, vol. 26, no. 1, pp. 242–256, 2020.

[6] Y. Xing, J. Yue, and C. Chen, "Dam deformation forecasting avoiding social and environmental impacts caused by dam break," *Fresenius Environmental Bulletin*, vol. 28, no. 11A, pp. 8831–8838, 2019.

[7] G. Maryam, O. Kaveh, E. Saeid, and P. Singh, "Application of time series modeling to study river water quality," *American Journal of Engineering and Applied Sciences*, vol. 11, no. 2, pp. 574–585, 2018.

[8] V. Grubinko, H. Humeniuk, V. Khomenchuk, N. Garmatiy, V. Voytiuk, and M. Barna, "Ecotoxicological status and prognosis of the state of an urbanized hydroecosystem (on the example of the reservoir "Ternopil pond")," *Journal of Geology, Geography and Geoecology*, vol. 27, no. 2, pp. 202–212, 2018.

[9] A. P. Belousova and E. E. Rudenko, "Basic environmental-hydrogeological studies in the territory of European Russia affected by chernobyl accident," *Water Resources*, vol. 46, no. 4, pp. 571–581, 2019.

[10] J. Xu, K. Wang, C. Lin, L. Xiao, X. Huang, and Y. Zhang, "FM-GRU: a time series prediction method for water quality based on seq2seq framework," *Water*, vol. 13, no. 8, pp. 1031–1045, 2021.

[11] T. Kato, A. Kobayashi, W. Oishi et al., "Sign-constrained linear regression for prediction of microbe concentration based on water quality datasets," *Journal of Water and Health*, vol. 17, no. 3, pp. 404–415, 2019.

[12] R. Jasim, M. Al-Saadi, and A. Khider, "New regression model for estimating irrigation water quality index," *International Journal of Design & Nature and Ecodynamics*, vol. 16, no. 2, pp. 127–134, 2021.

[13] H. Chiri, A. J. Abascal, S. Castanedo, and R. Medina, "Mid-long term oil spill forecast based on logistic regression modelling of met-ocean forcings," *Marine Pollution Bulletin*, vol. 146, no. Sep, pp. 962–976, 2019.

[14] C. Braun, "Parallel genetic algorithms for ground water pollution due to dumping of solid wastes," *Pollution Engineering*, vol. 1, no. 3, pp. 1–3, 2020.

[15] M. R. Enikeev, M. F. Fazlytdinov, L. V. Enikeeva, and I. M. Gubaidullin, "Forecast of water-cut at wells under design by machine learning methods," *Information Technology and Nanotechnology*, vol. 2019, no. 2416, pp. 510–520, 2019.

[16] A. S. Kalamdhad, J. Singh, and K. Dhamodharan, "Surface water quality modeling by regression analysis and artificial neural network," *Advances in Waste Management*, vol. 10, no. 15, pp. 215–230, 2019.

[17] M. S. Samsudin, A. Azid, S. I. Khalit, M. S. A. Sani, and F. Lananan, "Comparison of prediction model using spatial discriminant analysis for marine water quality index in mangrove estuarine zones," *Marine Pollution Bulletin*, vol. 141, no. 04, pp. 472–481, 2019.

[18] R. Barzegar, M. T. Aalami, and J. Adamowski, "Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model," *Stochastic Environmental Research and Risk Assessment*, vol. 34, no. 8, pp. 1–19, 2020.

[19] L. K. Narayanan, S. Sankaranarayanan, and J. Rodrigues, "CCC publications water demand forecasting using deep learning in IoT enabled water distribution network," *International Journal of Computers, Communications & Control*, vol. 15, no. 6, pp. 1–15, 2020.

[20] Z. Yang, H. Cai, W. Shao et al., "Clarifying the hydrological mechanisms and thresholds for rainfall-induced landslide: in situ monitoring of big data to unsaturated slope stability analysis," *Bulletin of Engineering Geology and the Environment*, vol. 78, no. 4, pp. 2139–2150, 2019.

[21] M. Azrour, J. Mabrouki, A. Guezzaz, and Y. Farhaoui, "New enhanced authentication protocol for Internet of things," *Big Data Mining and Analytics*, vol. 4, no. 1, pp. 1–9, 2021.

[22] H. Biglari, M. Tatari, M. R. Narooie, G. Ebrahimzadeh, and H. Sharafi, "Data for inactivation of free-living nematode rhabditida from water environment using ultraviolet radiation," *Data in Brief*, vol. 18, no. 4, pp. 30–34, 2018.

[23] N. M. Said, Z. M. Zin, M. N. Ismail, and T. A. Bakar, "Univariate water consumption time series prediction using deep learning in neural network (DLNN)," *International Journal of Advanced Technology and Engineering Exploration*, vol. 8, no. 76, pp. 473–483, 2021.

[24] A. Y. Ayturan, Z. C. Ayturan, and H. O. Altun, "Short-term prediction of PM2.5 pollution with deep learning methods," *Global Nest Journal*, vol. 22, no. 1, pp. 126–131, 2020.

[25] Z. Xie, Q. Liu, and Y. Cao, "Hybrid deep learning modeling for water level prediction in yangtze river," *Intelligent Automation & Soft Computing*, vol. 28, no. 1, pp. 153–166, 2021.

[26] C. Y. Hu, J. Y. Cai, Z. Zeng, X. S. Yan, W. Y. Gong, and L. Wang, "Deep reinforcement learning based valve scheduling for pollution isolation in water distribution network," *Mathematical Biosciences and Engineering: MBE*, vol. 17, no. 1, pp. 105–121, 2019.