Research Article

Edge Computing for Water Quality Monitoring Systems

Jianxun Ren, Qiliang Zhu, and Changsheng Wang

1 North China University of Water Resources and Electric Power, Zhengzhou, China
2 Water Resources Information Center of Henan Province, Zhengzhou, China
3 Xixiayuan Multipurpose Dam Water Diversion and Irrigation Project Construction and Management Bureau of Henan Province, Nanyang, China

Correspondence should be addressed to Qiliang Zhu; zhuqiliang@ncwu.edu.cn

Received 22 March 2022; Accepted 6 May 2022; Published 23 May 2022

Academic Editor: Yan Huang

Copyright © 2022 Jianxun Ren et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at the problems of insufficient real time and accuracy of water quality online monitoring and large resource consumption, this paper proposes a water quality monitoring and early warning method based on edge computing. Combined with Internet of things technology and edge computing technology, an online water quality monitoring and early warning model is designed. Through the preprocessing of the collected source data, the monitoring accuracy is improved, and edge computing technology is introduced to preliminarily analyze and process the collected data in the monitoring station, so as to save network traffic and computing resources. On the basis of online monitoring, the water quality prediction model is established by using historical water quality monitoring data to realize water quality prediction and provide a basis for staff scientific decision-making. Engineering practice shows that the model has high application value.

1. Introduction

With the rapid development of information technology, the Internet, Internet of things, cloud computing, 5G, and other technical means have been used in the collection, storage, calculation, and management of water conservancy information, which has improved the modernization level of water conservancy management ability. As a necessary means to detect the degree of water pollution and analyze the causes of water pollution, water quality monitoring is an important basis for water resources management and protection. Accurate monitoring of water resources with the help of modern emerging technologies is of great significance to improve the ecological environment [1]. In recent years, the automatic monitoring system composed of sensing technology, automatic measurement and control technology, computer technology, and communication network has achieved remarkable results [2]. However, there are still many problems in the actual production process of these technologies, which are as follows:

(i) Due to the error of sensor devices and the influence of environmental factors, there is more or less some noise in the source data collected by the water quality sensor, which is not conducive to post-processing and has an adverse impact on the accuracy of water quality monitoring.

(ii) Sensor devices based on the Internet of things generate a large amount of measurement data and transmit it to the cloud processor for analysis through the network, which will consume a lot of network traffic [3]. In addition, when the water quality is normal, the probability of data duplication is very high, so the network traffic consumed has no practical significance.

(iii) The large amount of redundant data generated by the measuring equipment brings a heavy burden to the later analysis and prediction, wastes computing resources, and has a negative impact on the real-time performance of the monitoring results.

(iv) Water quality monitoring can only reflect the current state of water quality and cannot make an effective judgment and early warning on the development trend of water quality.
The emergence of edge computing technology makes it possible to process small data generated by sensing devices in the Internet of things in real time locally [4]. This technology is a method of processing data close to the location of data generation. It adopts an open platform integrating network, computing, storage, and application core capabilities and emphasizes processing data nearby, so as to reduce the system response time, protect data privacy and security, prolong battery life, save network bandwidth, etc. [5]. At the same time, it also meets the basic needs of real-time business, application intelligence, security, and privacy protection. In the scene of water quality monitoring, although the amount of data generated by the water quality monitoring sensor equipment is small, it has high requirements for real-time processing. Therefore, the application of edge computing technology for water quality monitoring is of positive significance to improve the real-time monitoring and reduce the network flow.

In this paper, edge computing is applied in the field of water quality monitoring. Some computing tasks are unloaded to the data edge side for local processing, which can realize real time, reliable, and safe water quality monitoring and management.

2. Related Work

2.1. Water Quality Monitoring. Due to the increasingly serious water environment pollution, since the 1970s, the United States, Britain, Japan, the Netherlands, and Germany have successively established water quality monitoring systems. At present, the commonly used algorithm models mainly include the grey system model, support vector machine, multiple linear regression model, and artificial neural network [6]. Among them, the grey theory GM (1, 1) model has high requirements for the accuracy of historical data. If there is too much unknown information, the prediction error will increase and the stability of the model will deteriorate [7]. Although the multiple linear regression method has a simple principle and convenient modeling, it is mostly suitable for the application environment with the good linear condition [8]. As a very common algorithm for machine learning, SVM has the same function as a neural network and can solve many mathematical problems that cannot be solved by traditional methods. However, with the increase of data, the advantages of SVM will weaken. Due to the restriction of the Mercer condition, the selection of kernel function will be limited, which is only applicable to the modeling of small sample problems [9]. Although the neural network has the disadvantages of forgetfulness and difficulty to adjust the weight online, it has the ability to approach any nonlinear problem and can predict better on the whole.

In terms of platform design, many recent technologies use wireless sensor networks or the Internet of things as platforms for water quality monitoring and evaluation. For example, El-Deen et al. [10] proposed a low-cost wireless sensor network solution for real-time water quality monitoring, which has the advantages of low cost, lightweight, and self-organization. Tsai et al. [11] proposed a smart aquaculture system based on the Internet of things, which is used to detect the water quality of farms and provide automatic aeration to improve the survival rate of aquatic products. However, recently developed systems using wireless sensor network technology report deficiencies in energy management, data security, and communication coverage [12]. Although the Internet of things shows more efficient, safer, and cheaper advantages in application, the large amount of data generated by the Internet of things poses new challenges to data transmission and data processing.

2.2. Edge Computing. Edge computing is a research hotspot in recent years. It is a computing and network resource between the data source and cloud center, which can provide edge processing of big data [11]. In edge computing, the data will be calculated, stored, and applied at or near the Internet of things terminal. It is not necessary to upload all data to the cloud, which can effectively reduce the amount of communication transmission data. Massive data do not need centralized control decision-making. Using edge computing for distributed decision-making can reduce short-term delay and reduce user response time. Because the data are processed near the user side, it can also effectively avoid the risk of privacy disclosure caused by long-distance transmission.

Edge computing can effectively solve the problems of high latency, network instability, and low bandwidth in cloud computing. It has been applied to smart transportation, smart city, power grid detection, and other fields. In recent years, some scholars have applied edge computing technology to the field of water conservancy. Janet al. [12] combined edge computing and GIS technology to build a prediction model of ecological water demand. The model collects image data with GIS technology and divides different water resources in a timely and fast manner through edge computing data processing, so as to clarify the relationship between water resources and ecological environment. Abbas et al. [13] proposed a data link management solution based on mobile edge computing technology, which effectively realized the sinking of the service anchor and greatly shortened the service response time. Li et al. [14] proposed an incentive-based intelligent water-saving and distribution framework integrating blockchain and edge computing. The system integrates blockchain and water consumption prediction model into one framework to achieve the purpose of encouraging people to save water and prevent waste.

3. Water Quality Monitoring Model Based on Edge Computing

Aiming at the problems of insufficient real time and accuracy and high resource consumption in online water quality monitoring, this paper proposes a water quality monitoring and early warning method based on edge computing. Due to the limitation of the length of the article, we focus on the work content of the edge layer.
3.1. Framework of the Model. As shown in Figure 1, the method model is divided into five levels, from bottom to top, including the perception layer, edge layer, communication layer, cloud computing layer, and application layer.

(i) **Perception Layer.** Data collection is the basis of informatization, and the perception layer is the cornerstone of intelligent water quality monitoring and early warning. This layer includes various Internet of things devices for water quality detection, which are used to sense water temperature, PH value, chemical oxygen demand, suspended solids, conductivity, oxygen demand, and other information to provide data sources for the system.

(ii) **Edge Layer.** The edge layer can solve the problem of data processing nearby by deploying a large number of edge nodes. On the one hand, edge nodes receive, process, and forward effective data from sensing terminals and provide certain storage, computing, and decision-making capabilities. On the other hand, the edge node interacts with the cloud platform data, uploads the optimized necessary data to the cloud, receives the calculation rules and early warning model issued by the cloud platform, and realizes the real-time prediction and early warning of water quality data. This layer mainly includes data transmission unit, edge server, gateway, and other equipment. Its main functions are data transmission, data processing, data prediction, decision-making, calculation, and unloading according to the edge computing capacity. When the water quality tends to deteriorate, it sends an early warning and uploads the monitoring results to the cloud computing layer through the communication layer according to the time upload strategy.

(iii) **Communication Layer.** Information transmission is the premise of informatization, and the communication layer is the link of information sharing. This layer is responsible for the communication between the edge layer and the computing layer. The communication modes adopted between the two layers include NB-IoT, 4G, 5G, and Ethernet and communication protocols such as WebSocket, Modbus, and http/https are mainly used for data transmission. The reliable transmission of the communication layer is an important guarantee for connecting the functions of water quality monitoring layers.

(iv) **Cloud Computing Layer.** This layer is mainly composed of a cloud computing server, data server, and web server. The cloud computing server is mainly responsible for establishing a connection with the edge layer, keeping the data transmission channel smooth, and responsible for data analysis, model training, and complex event processing. The data storage server mainly provides data storage services to store the timing data and status data transmitted by the edge layer. The web server mainly provides data query interface, security authentication interface, and device control interface to provide services for the application layer.

(v) **Application Layer.** The application layer analyzes and arranges according to the interface provided by the webservice. According to different user roles, it displays to users through mobile app, web service,
and smart screen, displays water quality monitoring data in real time, and sends out an early warning in time. This aims to realize user operation business, which can improve the level of scientific decision-making and intelligent management.

3.2. Preprocess the Original Data. The function of data preprocessing is to eliminate the interference caused by equipment vibration, environmental change, and other factors in the data collection process of water quality sensors and reduce errors, so as to improve the accuracy of water quality monitoring and the accuracy of the prediction model. Data preprocessing mainly includes the removal of abnormal values of original data and the correction of abnormal values and provides basic data for subsequent water quality monitoring data analysis and prediction model establishment. In this paper, the quartile method is used to identify outliers.

The quartile method is to arrange all the data into an ascending sequence \( X = [x_1, x_2, \ldots, x_{n-1}, x_n] \), where \( x_{(i)} < x_{(i+1)} \), \( n \) is the total number of samples, and \( x_i \) represents a point in the sequence. Sequence \( X \) is divided into four equal parts, in which each part contains 25% data. The dividing points of each data are the lower quartile \( q_1 \), median \( q_2 \), and upper quartile \( q_3 \), respectively. IQR is called interquartile interval, which is the difference between the upper quartile \( q_3 \) and the lower quartile \( q_1 \). The data in IQR account for half of all the data in the sequence. \( q_1, q_2, q_3 \) are calculated as follows:

\[
q_2 = \begin{cases} 
  \frac{x_{n+1/2}}{2} & n = 2k + 1; k = 0, 1, 2, \ldots \\
  \frac{x_{n/2} + x_{n+2/2}}{2} & n = 2k; k = 0, 1, 2, \ldots 
\end{cases}
\]

When \( n = 2k \) \((k = 1, 2, \ldots)\), divide \( X \) into two parts from \( q_2 \), where \( q_2 \) is not included in the two parts, and calculate the median \( q_2^1 \) and \( q_2^2 \) of the two parts, respectively, and then, \( q_1 = q_2^1, q_3 = q_2^2 \).

When \( n = 4k + 3 \) \((k = 1, 2, \ldots)\), then

\[
\begin{align*}
q_1 &= \frac{3}{4} x_{k+1} + \frac{1}{4} x_{k+2}, \\
q_3 &= \frac{1}{4} x_{3k+2} + \frac{3}{4} x_{3k+3}
\end{align*}
\]

When \( n = 4k + 1 \) \((k = 1, 2, \ldots)\), then

\[
\begin{align*}
q_1 &= \frac{1}{4} x_k + \frac{3}{4} x_{k+1}, \\
q_3 &= \frac{3}{4} x_{3k+1} + \frac{1}{4} x_{3k+2}
\end{align*}
\]

Finally, the interquartile distance can be calculated as follows:

\[
\text{IQR} = q_3 - q_1.
\]

The limits of outliers in the data sample are as follows:

\[
[F_1, F_u] = [q_1 - 1.5\text{IQR}, q_3 + 1.5\text{IQR}].
\]

If \( x \notin [F_1, F_u] \), determine \( x \) as abnormal data; on the contrary, it is determined that \( x \) is normal.

In this paper, the polynomial fitting method is used for the secondary identification and elimination of outliers. After the two eliminations of outliers, the outliers can be regarded as missing values, so the correction of outliers is to fill in the missing values. Therefore, the polynomial fitting of the sequence after eliminating outliers twice and filling the missing values with the fitted corresponding values can achieve the purpose of correction. The goal of the polynomial fitting is to minimize the sum of squares of errors. Suppose a function combination

\[
f(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \cdots + a_m\varphi_m(x).
\]

The sum of squares of errors is expressed as

\[
\|\delta\|^2 = \sum_{i=1}^{n} \delta_i^2 = \sum_{i=1}^{n} [f(x_i) - y_i]^2,
\]

where \((x_i, y_i) (i = 0, 1, \ldots, n)\) is a given set of data. Polynomial fitting is to find a curve that is closest to all data points on the premise of minimizing the sum of squares of errors, that is, to find \( f(x) \) that minimizes \( \|\delta\|^2 \).

The process of finding the fitting curve function is transformed into the problem of finding the minimum value of the multivariate function \( H(a_0, a_1, \ldots, a_m) \), and the multivariate function is expressed as follows:

\[
H(a_0, a_1, \ldots, a_m) = \sum_{i=1}^{n} \delta_i^2 = \sum_{i=1}^{n} [f(x_i) - y_i]^2.
\]

By solving the minimum value of the multivariate function, the solution \((a_0^*, a_1^*, \ldots, a_m^*)\) can be obtained, so that the least square solution of the function \( f(x) \) is

\[
f(x) = a_0^*\varphi_0(x) + a_1^*\varphi_1(x) + \cdots + a_m^*\varphi_m(x).
\]

Generally, the fitting degree of the polynomial fitting is 3 times. If it is less than 3 times, the peak of the curve may be lost; if it is higher than 3 times, the fitting time is too long, and it is easy to produce false peaks. In this paper, the fitting degree of the polynomial fitting is set to 3.

3.3. Task Migration and Data Transmission. Different from the traditional method of transmitting the collected data to the cloud for analysis, this topic uses the edge computing technology to analyze and process the collected data at the near end. After preprocessing the data collected by the perception layer, data calculation, state recognition, result transmission, and data prediction are carried out in the edge layer. When the processing capacity of the edge layer is insufficient, the computing task will be unloaded to the cloud. Figure 2 illustrates the process of task unloading and data transmission.

(i) Task Migration. The edge layer processes the data according to its own processing capacity. If its processing capacity is sufficient, the data processing will be carried out in the edge layer. On the contrary,
it will send the collected data to the cloud for processing. In the process of operation, we judge whether task migration is necessary according to the time of data processing. Suppose that the processing time of the task is $T_p$ and the threshold of the processing time is $T$. If $(T_p) > T$, it means that the task processing times out, and the computing task will be unloaded to the cloud for execution.

(ii) Abnormal Submission. If the calculation task is within the capability of the edge layer, the water quality detection results will be completed in the edge layer. Assuming that $R$ is a water quality detection result, $[R_{\min}, R_{\max}]$ represents the normal range of the result. If $r \notin [R_{\min}, R_{\max}]$, it indicates that the water quality is abnormal, and the water quality detection data will be transmitted to the cloud for further calculation and confirmation. If it is indeed an abnormal water quality, the early warning system is started.

(iii) Result Upload. If $r \in [R_{\min}, R_{\max}]$, the system will judge whether a time period has been reached. Suppose $t$ represents a result transmission time period, and $t_r$ represents the time of uploading the result for the $n$th time. If $t_{n+1} = t_n + t$, then the water quality detection results are uploaded to the cloud. Otherwise, the results will not be uploaded, but the next round of detection results will be calculated until a new upload cycle is reached. This way of periodically uploading results can not only ensure the consistency of detection data but also greatly save bandwidth and reduce data redundancy.

In short, once the processing capacity of the edge layer is insufficient, the computing task will be unloaded to the cloud for execution. When the water quality is normal, it is transmitted to the remote end according to the set frequency. When the water quality data are abnormal, it will be transmitted to the cloud in real time. This mode will greatly save bandwidth and ensure the real-time upload of detection results.

3.4. Water Quality Prediction. Water quality prediction is an important module of the model. Due to the variety and complexity of water environmental factors and the complex nonlinear relationship, many experts have focused on nonmechanistic water quality models and made some progress. However, many of these models have the disadvantage of large error. Considering that the artificial neural network model has strong adaptability, self-learning, and high fault tolerance, this paper uses BP neural network to predict water quality. BP neural network algorithm is a multilayer feedforward network trained by error backpropagation algorithm. As shown in Figure 3, the topology of a multilayer BP neural network consists of three layers: input layer, output layer, and hidden layer, in which there can be more than or equal to one hidden layer. There are several neurons in each layer. The neurons in the same layer have no relationship with each other, but only have input-output relationship with the neurons in adjacent layers. In order to improve the global convergence of the BP neural network, we use the hybrid optimization method based on the Nelder–Mead simplex method and cuckoo search algorithm to optimize the weight and deviation of the BP network. For details, refer to our previous work [15].

3.4.1. Determination of Input and Output Layer. The indicators used for water quality evaluation are water temperature, PH value, chemical oxygen demand, suspended solids, conductivity, oxygen demand, etc. In the water quality prediction model, the data of six indicators are normalized as the model input nodes and the water quality grade as the model output node. Therefore, the number of input layer nodes of the BP neural network is 6, and the number of output layer nodes is 1.

3.4.2. Hidden Layer Determination. According to Kolmogorov’s theorem and the least square approximation theorem of mapping, when the number of neurons is enough, the BP neural network with hidden layers of 1 can approximate any nonlinear function [16]. Therefore, the hidden layers of the BP neural network prediction model are determined as 1.

Determining the number of hidden layer nodes is a very important step in the process of initializing the network structure. Too many nodes in the hidden layer will increase the amount of calculation of the BP neural network and easily lead to the overfitting problem; if the number of hidden layer nodes is too small, it will affect the performance of network training and fail to achieve the expected effect. In order to improve the speed of network learning, this paper determines the best number of hidden layer nodes by comparing the error of verification set under different number of nodes.
3.4.3. Selection of Excitation Function. Sigmoid function is one of the most commonly used activation functions in the BP neural network, which is closest to biological neurons in the physical sense. It can compress a real value to the range of $[0, 1]$ and can keep the data amplitude from large changes. The function of the sigmoid function can meet the application requirements of the solution in this paper. Therefore, the sigmoid function is selected as the activation function in this paper. The detailed description is as follows:

$$f(x) = \frac{1}{1 + e^{-x}}.$$ (9)

When $x$ approaches negative infinity, $f(x)$ approaches 0; when $x$ approaches positive infinity, $f(x)$ approaches 1; when $x = 0$, $y = 1/2$.

3.4.4. Weight and Deviation Initialization. The weight threshold initialization of the BP neural network is usually realized by generating random numbers. In this paper, the weights and deviation random numbers between (0, 1) are randomly generated, and then, the hybrid optimization algorithm based on the Nelder–Mead simplex method and the cuckoo search algorithm is used to obtain the final initialization weights and deviation values.

An improved neural network algorithm is proposed to establish a water quality prediction model. Using the preprocessed historical data to train the prediction model, get the optimal parameters of the model, verify the trained model, and evaluate and analyze the prediction results.

4. Engineering Application

This paper takes the water conveyance and irrigation project of the Xixiayuan water control project in Henan Province as the research object. The function of the Xixiayuan project is mainly reverse regulation, combined with power generation and comprehensive utilization of water supply and irrigation. While ensuring the continuous flow of the Yellow River, it fundamentally eliminates the adverse impact of peak shaving of Xiaolangdi Hydropower Station on the downstream river and plays a vital role in ecological, environmental protection, and industrial and agricultural production water.

Experimenters deployed water quality detection sensors and edge gateways in the canal. The sensor is connected to the edge gateway through RS485 serial port line. The edge gateway can upload data to the cloud through mobile communication, and the content of the cloud platform can be viewed in real time through the web. The cloud platform sets water quality early warning rules and triggers early warning information according to the data uploaded by edge nodes. Table 1 shows the average data transmission response time of different deployment schemes. We can see that the average response time of the cloud side system computing scheme proposed in this paper is 36 ms, which is basically consistent with that of the local computing scheme, but far lower than that of the cloud computing mode. It shows that the model proposed in this paper makes a significant contribution to the real-time performance of data transmission. In terms of water quality prediction, the accuracy of using the BP neural network based on the original collected data is 88.92%, which is much higher than that of the multiple linear regression prediction method. The accuracy

<table>
<thead>
<tr>
<th>Deployment scheme</th>
<th>Average response time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local computing</td>
<td>33</td>
</tr>
<tr>
<td>Cloud-edge computing</td>
<td>36</td>
</tr>
<tr>
<td>Cloud computing</td>
<td>113</td>
</tr>
</tbody>
</table>

![Figure 3: Topology of three-layer BP neural network.](image)
of using the BP neural network based on the preprocessed data is 91.16%. Obviously, the prediction model and data preprocessing method proposed in this paper have also achieved ideal results.

5. Conclusion
In order to improve the real-time transmission of water quality monitoring data and the accuracy of water quality early warning, a water quality monitoring and early warning model based on edge computing is proposed in this paper. The model makes full use of the computing power of the edge layer to process the water quality detection data collected by the perception layer. According to the actual needs, this paper designs the task migration and data transmission rules of the edge layer and puts forward the methods of data preprocessing and water quality prediction. Engineering practice shows that the model and method proposed in this paper have high application value. In future research, we will pay more attention to the research of water quality early warning and the application of edge computing in safety supervision [17, 18].

Data Availability
The data used to support the findings of this study have been deposited in the Baiduyun (https://pan.baidu.com/s/117-G1NR3TVXC9gWjmmMaFQ?pwd=m000).

Conflicts of Interest
The authors declare do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Acknowledgments
This work was supported by the high-level talents research initiation project of North China University of Water Resources and Electric Power (No. 201811026) and the Science and Technology Project of Henan Province (No. 222102240010).

References