Research on water supply network pressure prediction using PSO-BP neural network algorithm

Chaoran Wang*, Changtao Wang, Dan Shan, Baolong Yuan
School of Automation and Electrical Engineering, Shenyang Jianzhu University, Shenyang 110168, Liaoning, China

ABSTRACT

Ensuring the accurate prediction of water supply network pressure is crucial for the efficient operation of the water supply system. The purpose is to address the issue of insufficient accuracy in short-term forecasting. The method of short-term prediction for network pressure nodes using the Particle Swarm Optimization-Back Propagation (PSO-BP) neural network algorithm is proposed in this paper. Firstly, the real pressure data collected from a water supply network, upon which this paper relies, is cleaned using the Cook's distance method. After the dirty data is removed, Back Propagation (BP), Genetic Algorithm-Back Propagation (GA-BP), and PSO-BP are employed to predict the pressure in the water supply network. By comparing the prediction results, PSO-BP was found to be the most accurate among the three algorithms, with an RMSE (Root Mean Square Error) of 1.9132. In order to solve the problem of insensitive regions in pressure prediction, a variable sliding window method is proposed to determine the data set based on the previous method. The results indicate that this method can effectively improve the accuracy of prediction in insensitive areas.

Keywords: Water supply network, network pressure prediction, PSO-BP algorithm, data cleaning

1. INTRODUCTION

With the acceleration of urbanization, the expansion of urban population and industrial production scale has led to a sharp increase in the demand for water supply. This requires the water supply system to provide a stable and sufficient amount of water with appropriate water pressure to meet the needs of residents' daily life and industrial production. Therefore, accurate prediction of water supply network pressure is crucial for ensuring efficient operation and meeting demand in the water supply system. Pressure prediction can provide a foundation for other network optimizations, such as leak detection and network scheduling. Currently, issues related to urban water supply network pressure are receiving significant attention from a large number of researchers. Dawood et al.[1] have achieved certain results in the assessment of water supply pipe leakage by employing machine learning models. Letting et al.[2] applied an intelligent stochastic search algorithm to the calibration of hydraulic model parameters in water supply networks. Viccione et al.[3] applied the ARIMA model for water level simulation and prediction. Ye et al.[4] utilized the linear Kalman filter and the linear unbiased minimum mean square error estimation criterion to predict the pressure or flow of the water supply network at the current moment. Nerantzis et al.[5] utilized model predictive control to optimize the scheduling of pipeline networks, which involves using a pipeline network model in conjunction with optimization algorithms to select the best scheduling plan. Agafonov et al.[6] employ Recurrent Neural Network (RNN) to analyze time series data and have achieved outstanding results. According to Ismail et al.[7], accelerometer sensors should be applied to detect vibrations from water pipeline leaks. Mamo et al.[8] proposed a leakage detection and classification method based on multi-class support vector. Jesús et al.[9] constructed a leakage prediction model for water supply networks using topology differential evolution. Through reviewing relevant literature and collecting and organizing related calculations, the author conducted predictions on actual pipeline networks using Matlab. The author processed and cleaned the dataset using the Cook's distance method, achieving a higher accuracy rate. Based on the insensitive points identified in the predictions, the author proposes a new method, the variable sliding window method, which aims to provide a new perspective for solving the problem of pipeline pressure prediction in the future.

*578980829@qq.com

International Conference on Optics, Electronics, and Communication Engineering (OECE 2024), edited by Yang Yue, Proc. of SPIE Vol. 13395, 1339532 · © 2024 SPIE · 0277-786X · Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.3049406

2. METHODOLOGY

2.1 The significance of data cleaning

Data anomalies in water supply network monitoring refer to abnormal conditions such as missing data values, incorrect data records, and other irregularities found in the monitoring records of the water supply network. An essential prerequisite for carrying out data cleaning in water supply network monitoring is to organize and summarize the anomalies in these data, and to conduct research and analysis on the patterns, categories, quantities, and causes of data anomalies. Based on the analysis and processing of a large amount of monitoring data, this article describes and classifies the anomalies in the monitoring data across different dimensions, serving as the foundation for data cleaning and quality assessment.

Data cleaning involves two parts of work. The first is to identify and detect errors in the data, which is also known as anomaly detection or anomaly identification, such as discovering data loss or identifying outliers, etc. The second part involves correcting and repairing the identified anomalies, such as replacing outliers, filling in missing data, etc. Data cleaning is a foundational task that "connects the previous and subsequent steps" based on the analysis of existing data, aiming to meet the data quality requirements of subsequent business operations.

2.2 Cook's distance

After performing regression prediction, if the accuracy of the results is not satisfactory, data cleaning can indeed have an impact on the accuracy of the regression prediction. The data was analyzed and processed using IBM-SPSS software. This study selected Cook's distance as a method to identify outliers. Before data cleaning, the entire dataset is first analyzed and graphed to observe the distribution of pressure points over time across 31 days, facilitating subsequent comparisons and analyses, as shown in Figure 1.

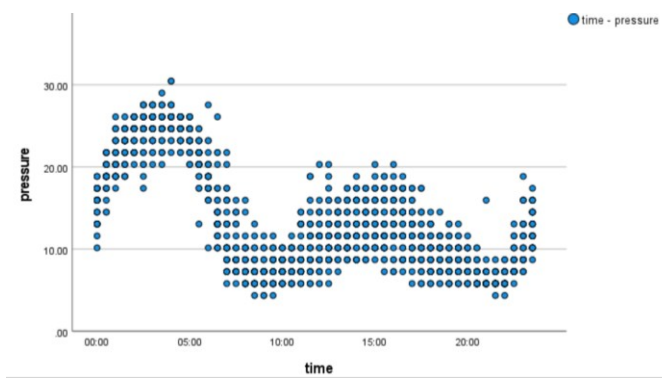

Figure 1. Time pressure scatter plot (original).

This article continues to discuss the analysis of outliers in the data using Cook's distance as the method. Extract all days from August 2023 for individual analysis, with one day's image shown in Figure 2.

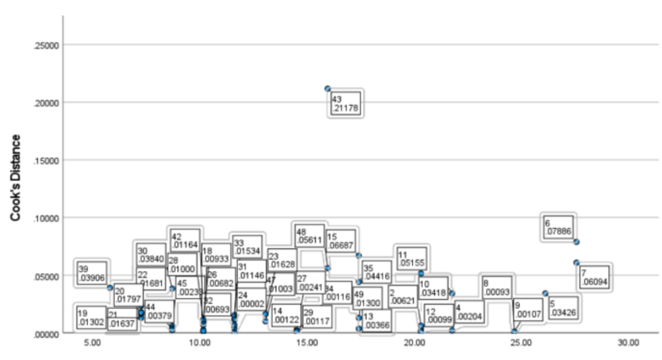

Figure 2. Pressure cook's distance scatter plot.

From the annotations on the graph above, it can be deduced that points 48 and 14 are outliers. Extending this logic, this article has analyzed all 31 days in August and aggregated the results into the dataset for further analysis. Based on the above conclusions, the dataset was reanalyzed by replacing the outliers with the average value of the specific time on the 31st day. After reanalyzing the entire dataset, Figure 3 was obtained. By comparing the two graphs, it can be observed intuitively that the outliers have been removed, and the entire curve fits more closely.

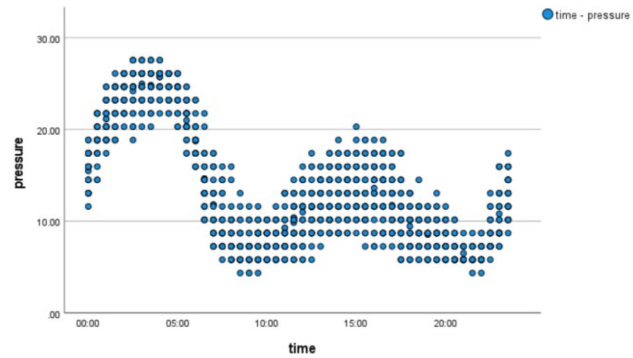

Figure 3. Time pressure scatter plot (cleaned).

2.3 PSO-BP regression prediction

The PSO-BP algorithm refers to a neural network training method that combines Particle Swarm Optimization (PSO) with Backpropagation (BP). Here is a sample illustration and caption for a multimedia file. This algorithm combines the global search capability of the Particle Swarm Optimization (PSO) algorithm with the local search capability of the Backpropagation (BP) algorithm, aiming to improve the training efficiency and performance of neural networks. Using the PSO algorithm to optimize the weights and biases of the BP neural network can enhance the training efficiency and performance of the network. Here are the basic steps for optimizing a BP neural network using PSO:

Step 1: Particle Swarm Initialization: The positions of the particles, which represent the weights and biases of the BP neural network, should be randomly initialized. The velocity of the particles should also be randomly initialized. The size of the particle swarm, which is the number of particles, needs to be set.

Step 2: Particles Evaluation: The fitness of each particle should be evaluated using the error function of the BP neural network. The closer the network weights and biases represented by the particle are to the optimal solution, the smaller the value of the fitness function will be.

Step 3: Individual Best and Global Best Updating: For each particle, the current fitness should be compared with the historical best fitness. If the current fitness is better, the individual best position of the particle should be updated. The global best fitness among all particles should be found, and the corresponding particle position should be recorded.

Step 4: The velocity and position of the particles should be updated.

Step 5: Termination Condition Check: If the maximum number of iterations is reached or the fitness reaches a predetermined threshold, the iteration should be stopped. Otherwise, the process should return to Step 2 to continue the iteration.

Step 6: Global Optimal Solution Output: Upon termination of the algorithm, the global optimal position should be outputted, which represents the optimal weights and biases for the BP neural network.

2.4 Sliding window regression prediction

This paper introduces a variable sliding window approach, which applies a more detailed sub-prediction method using time-series forecasting when predicting insensitive areas. The program will predict in 5-minute increments, using a 35-minute cycle to forecast specific fluctuation points. When the sliding window prediction with a daily cycle identifies points with significant fluctuations, as shown in Figure 4.

Sub-predictions are conducted using a variable sliding window, and a 35-minute cycle can cover the previous point of the fluctuation point, with the previous point being a normal value. Therefore, it can provide more detailed changes in pressure points at 5-minute intervals to predict the situation at that point. Moreover, this time-series prediction can make more accurate predictions for fluctuation areas, which not only improves the overall prediction accuracy but also allows for analysis of abnormal fluctuation points.

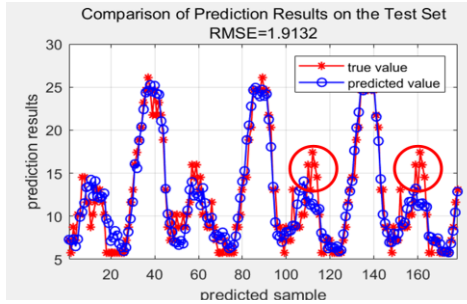

Figure 4. PSO-BP test set prediction of fluctuation range.

3. DATA

Data collection is the foundation of data stream analysis, and through the Internet, various sensors, and monitoring equipment, real-time monitoring and data acquisition of flow, pressure, water quality, and other data in the water supply system is conducted.[10] By collecting the actual operation conditions of the pipeline network in a certain area of a city over the past three years as a dataset, and through data collection and processing, the model was constructed and validated for all pressure data at the site in August 2023. When modeling in Matlab, the 10-fold cross-validation method is used to divide the dataset into a training set and a test set at a ratio of 9:1. This design employs the sliding window method to construct the model, which can break down long sequences into short ones, thereby reducing computational load and memory usage. Each subsequence can be processed independently, enhancing processing efficiency. Suitable for large-scale datasets and high-dimensional data; can be combined with other algorithms for processing, with strong scalability. In the experiment, we use data from the previous seven days to predict the next day, so the window size is set to 7 and the step size is set to 1. It is shown in Figure 5. In the experiment, based on a daily basis, we further refine the data by dividing each day into 30-minute intervals to predict the pressure value every 30 minutes throughout the day. It is shown in Figure 6. The data after final data cleaning is displayed in Figure 7.

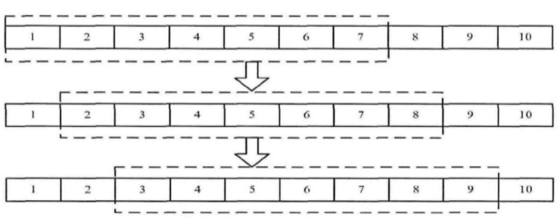

Figure 5. Implementation diagram of the sliding window method.

0:30:00	17.40	18.85	18.85	18.85	20.30	20.30	15.95	21.75
1:00:00	18.85	20.30	20.30	20.30	20.30	24.65	21.75	21.75
1:30:00	18.85	21.75	23.20	23.20	21.75	24.65	23.20	24.65
2:00:00	21.75	21.75	21.75	21.75	26.10	24.65	23.20	26.10
2:30:00	20.30	23.20	24.65	24.65	27.55	27.55	18.85	24.65
3:00:00	23.20	23.20	26.10	26.10	27.55	26.10	23.20	26.10
3:30:00	23.20	24.65	26.10	24.65	24.65	27.55	21.75	26.10
4:00:00	26.10	27.55	26.10	27.55	24.65	26.10	23.20	26.10
4:30:00	26.10	24.65	26.10	26.10	21.75	26.10	23.20	24.65
5:00:00	26.10	24.65	23.20	24.65	20.30	24.65	20.30	23.20
5:30:00	18.85	20.30	23.20	20.30	20.30	24.65	13.05	17.40
6:00:00	17.40	18.85	18.85	20.30	17.40	18.85	10.15	15.95
6:30:00	13.05	17.40	13.05	21.75	14.50	17.40	10.15	11.60

Figure 6. Partial data before data cleaning.

0:30:00	17.40	18.85	18.85	18.85	20.30	20.30	15.95	21.75
1:00:00	18.85	20.30	20.30	20.30	20.30	24.65	21.75	21.75
1:30:00	18.85	21.75	23.20	23.20	21.75	24.65	23.20	24.65
2:00:00	21.75	21.75	21.75	21.75	26.10	24.65	23.20	26.10
2:30:00	20.30	23.20	24.65	24.65	27.55	27.55	18.85	24.65
3:00:00	23.20	23.20	26.10	26.10	27.55	26.10	23.20	26.10
3:30:00	23.20	24.65	26.10	24.65	24.65	27.55	21.75	26.10
4:00:00	26.10	25.73	26.10	27.55	24.65	26.10	23.20	26.10
4:30:00	26.10	24.65	26.10	26.10	21.75	26.10	23.20	24.65
5:00:00	26.10	24.65	23.20	24.65	20.30	24.65	20.30	23.20
5:30:00	18.85	20.30	23.20	20.30	20.30	24.65	21.00	17.40
6:00:00	17.40	18.85	18.85	20.30	17.40	18.85	18.75	15.95
6:30:00	13.05	17.40	13.05	14.64	14.50	17.40	10.15	11.60

Figure 7. Partial data after data cleaning.

4. RESULTS

4.1 Comparison of prediction results before and after cleaning

To verify the impact of the above data analysis on prediction accuracy after cleaning, this study will use PSO-BP to predict the cleaned dataset separately. The prediction results before data cleaning are shown in Figure 8.

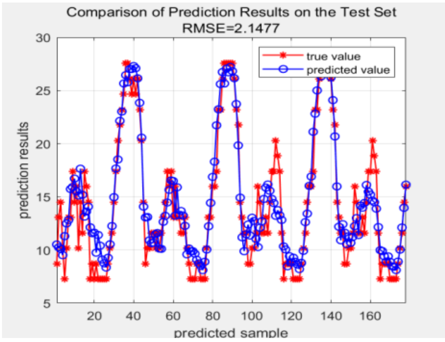

Figure 8. The prediction results before data cleaning.

The results after data cleaning are shown in Figure 9.

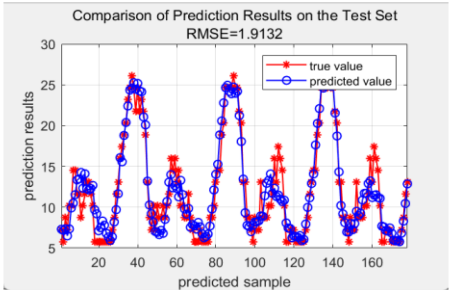

Figure 9. The prediction results after data cleaning.

As can be seen, after data cleaning, there is a significant change in the RMSE (Root Mean Square Error), indicating that the data processed by cleaning with Cook's distance has greatly improved the accuracy of regression predictions.The specific results can be seen in Table 1.

Table 1. Comparison of RMSE before and after cleaning.

	Before cleaning	**After cleaning**
RMSE	2.1477	1.9132

4.2 Comparison of multiple algorithms

Based on the PSO-BP algorithm, the intention is to verify how the regression prediction accuracy of PSO-BP compares to other algorithms. This article also selected two other algorithms, namely BP neural network and GA-BP. These two algorithms were also used for regression prediction on this pressure node. The results obtained were compared to draw conclusions.

Firstly, the BP neural network was used, resulting in Figure 10.

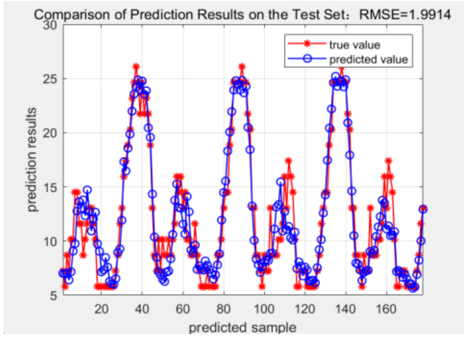

Figure 10. BP test set prediction chart.

GA-BP algorithm, which stands for Genetic Algorithm Optimized Backpropagation Neural Network, is a hybrid optimization method that combines Genetic Algorithm (GA) with Backpropagation (BP) neural networks. This algorithm aims to optimize the parameters of the BP neural network using the global search capability of the Genetic Algorithm (GA), thereby enhancing the prediction accuracy and performance of the network.

Secondly, the results of GA-BP are shown in Figure 11.

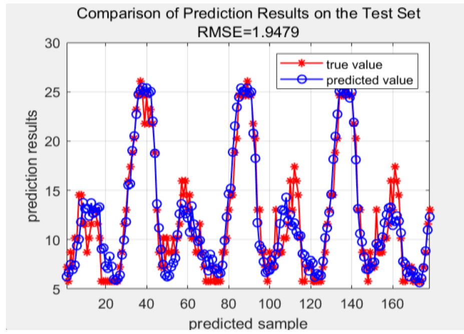

Figure 11. GA-BP test set prediction chart.

The results obtained by synthesizing the three types are shown in Table 2.

Table 2. Comparison of RMSE for BP, GA-BP, and PSO-BP.

	RMSE
BP	1.9914
GA-BP	1.9479
PSO-BP	1.9132

It can be seen that the accuracy of PSO-BP is higher compared to other algorithms.

4.3 Comparison of variable sliding window regression prediction

Using PSO-BP for variable sliding window regression prediction resulted in Figure 12.

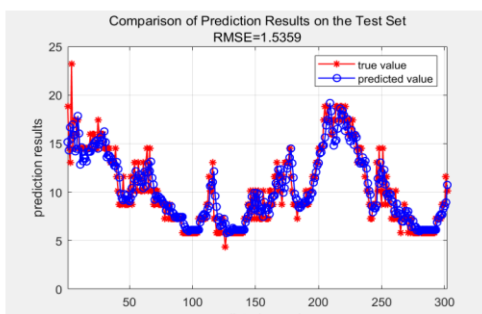

Figure 12. PSO-BP variable sliding window test set prediction chart.

The results of the sliding window method and the variable sliding window method are shown in Table 3.

Table 3. Comparison of RMSE for the sliding window method and the variable sliding window method.

	Sliding window method	**Variable sliding window method**
RMSE	1.9132	1.5359

Through the comparison of the above results, it was found that the variable sliding window method proposed in this paper provides more accurate prediction results in the insensitive areas of the sliding window method.

5. CONCLUSIONS

PSO-BP is employed in this paper to predict the pressure of a single-node operating condition in a specific city. The Cook's distance method is used to identify outliers in the dataset, and to label and modify the relevant data within the dataset. After using the cleaned dataset for prediction, more accurate prediction results were obtained with an RMSE=1.9132. To verify whether PSO-BP provides more accurate pressure predictions compared to other algorithms, the dataset was cleaned and the results of predictions using BP, GA-BP, and PSO-BP algorithms were compared. It was found that PSO-BP achieved a higher accuracy rate. RMSE=1.9132<1.9749<1.9914. Building upon the sliding window approach, a method of variable sliding window is proposed for a more detailed study to address the issue of inaccurate predictions in insensitive areas when using PSO-BP with a sliding window. By applying the variable sliding window method, more accurate results were obtained in the insensitive areas. A better result obtained is RMSE=1.5359<1.9132. The above methods and experiments provide more insights for studying water supply network pressure prediction.

REFERENCES

[1] Dawood, T., Elwakil, E., Novoa, H. M., et al., "Water pipe failure prediction and risk models: State-of-the-art review," Canadian Journal of Civil Engineering 47(10), 1117-1127 (2020).

[2] Letting, L. K., Hamam, Y. and Abu-Mahfouz, A. M., "Estimation of water demand in water distribution systems using particle swarm optimization," Water 9(8), 593 (2017).

[3] Viccione, G., Guarnaccia, C., Mancini, S., et al., "On the use of ARIMA models for short-term water tank levels forecasting," Water Supply 20(3), 1-13 (2020).

[4] Yeg Fenner, A. R., "Kalman filtering of hydraulic measurements for burst detection in water distribution systems," Journal of Pipeline Systems Engineering and Practice 2(1), 14-22 (2011).

[5] Nerantzis, D. and Stoianov, I., "Adaptive model predictive control for fire incidents in water distribution networks," Journal of Water Resources Planning and Management 148(2), 04021102 (2022).

[6] Agafonov, A. A. and Yumaganov, A. S., "Bus arrival time prediction using recurrent neural network with LSTM architecture," Optical Memory and Neural Networks 28(3), 222-230 (2019).

[7] Ismail, M., Dziyauddin, R A., Salleh, N. A. A., et al., "A review of vibration detection methods using accelerometer sensors for water pipeline leakage," IEEE access7 (18), 51965-51981 (2019).

[8] Mamo, T. G., Juran, I. and Shahrour, I., "Virtual DMA municipal water supply pipeline leak detection and classification using advance pattern recognizer multi-class SVM," Journal of Pattern Recognition Research 9(1), 25-42 (2014).

[9] Jesús, M. A, Marcos, Q., Cristina, V., et al., "A leak zone location approach in water distribution networks combining data-driven and model-based methods," Water 13(20), 2924-2924 (2021).

[10] Oberascher, M., Rauch, W. and Sitzenfrei, R., "Towards a smart water city: A comprehensive review of applications, data requirements, and communication technologies for integrated management," Sustainable Cities and Society 76, 103442 (2022).