Energy Proceedings
Vol 26, 2022
ISSN 2004-2965

Pipeline Block Localization in Water Distribution Networks Using Artificial Neural Network#

Dinghuang Xing[1], Guang Yang[1], Hai Wang[1, *]

1 School of Mechanical Engineering, Tongji University, Shanghai, China200092

* Correspondence: wanghai@tongji.edu.cn

ABSTRACT

The pipeline blockage increases the resistance of water distribution networks. In this study, a new method based on machine learning was proposed to locate the blocked pipeline. Numerous block scenarios were simulated by the hydraulic simulation, considering various block sizes and user demands for each pipeline. The dataset of the pressure change rates on the nodes was used to train artificial neural network models. The influences of the dataset variables on the model performance were analyzed. Results showed that the proposed method can successfully locate the blocked pipeline using the measurement of one day.

Keywords: water distribution networks, artificial neural network, block localization

NONMENCLATURE

Abbreviations	
WDNs	Water Distribution Networks
SVM	Support Vector Machine
kNN	k-Nearest-Neighbor
ANN	Artificial Neural Network
Symbols	
C_d	Diameter Coefficient

1 INTRODUCTION

Water distribution networks (WDNs) are the infrastructures of cities and maintain the operation of the water supply. Clogging in the pipelines of WDNs may occur. The causes of blockages in the pipelines can be chemical corrosion, deposition of minerals, partial closure of valves, cold icing, etc. When the blockage exists in a pipeline, the high resistance of the pipeline increases the upstream pressure and the risk of water burst. And the downstream flow rate decreases, which may fail the demands of end users. Therefore, it is necessary to locate the block failure of WDNs.

Blockage in the pipeline can be classified as discrete blockage and extended blockage depending on the length of the blocked portion relative to the pipeline [1]. The blockage affects the flow pattern of water and thus signal anomalies can be detected by sensors. To find the relationship between the abnormal signals and the blockage in the pipeline, the signal analysis method has been used to detect the blockage. The signals collected are in various forms, such as vibration signal, acoustic signal and pressure signal. Lile et al. [2] collected vibration signals by accelerometers and found that the vibration was stronger when the flow area was smaller. Yang et al. [3] proposed a signal noise reduction method based CEEMD-VT-SVD to improve the performance of analyzing acoustic signals to identify the blockage. Sattar et al. [4] used pressure sensors to collect the pressure signals at the end of the pipeline and analyzed the effect of the discrete blockage in the pipeline by using the frequency response method. Duan et al. [5] studied the effect of the extended blockage in the pipeline on the pressure signal using the frequency response method. Kim et al. [6] proposed a simplified formula to represent multiple discrete blockages in the pipeline. Lee et al. [7] used the frequency response diagram to detect single and multiple partial blockages in the pipeline. Massari et al. [8, 9] used the stochastic successive linear estimation method to establish the relationship between the pressure signal and the pipeline diameter to estimate the size and the location of the blockage.

The signal analysis method shows its potential for detecting the blockage in the pipeline. However, for urban water supply networks, numerous pipelines are interconnected to form a complex network. The varying demands of end users unsteady the flow state in the network, which brings perturbations to the signal analysis method making analyzing difficult. Recently, the machine learning method has been received attention from researchers due to its powerful characterization capability. The machine learning method obtains data from pressure and flow sensors in WDNs, and builds a machine learning model to identify the fault data which indicates the fault location in WDNs. At present, to the authors' knowledge, no research has been reported on the localization of pipeline blocks in WDNs with the machine learning method. As with blocks, leaks in WDNs affect the flow state, influencing the data measured from pressure and flow sensors. Hence, leak localization methods in WDNs should be reviewed. Mashford et al. [10] used support vector machine (SVM) to predict the size of a single leak and another SVM to predict the leak location with measured pressures from six points, and 97.7% of the predicted nodes were within 300 meters of the real nodes. Wachla et al. [11] used SVM to build flow prediction models for 23 zones dividing by a large WDN and used ANFIS to determine whether a leak occurs for each zone. Zhang et al. [12] divided a large WDN into multiple subzones using the k-means method and used SVM to predict which subzone the leak occurred in. The performance of SVM model was lower when more subzones were divided. This method narrows the leak events into subzones, which is helpful for on-side leak localization.

Further studies investigated the effects of uncertainties on the model performance of leak localization in WDNs. Soldevila et al. [13] used Bayesian classifier to predict the location of a single leak in the WDN. The comparison was conducted on the performances of Bayesian classifier, k-Nearest-Neighbor (kNN) classifier and the angle method in four cases of uncertainties: leak size, pressure measurement, user demand, and the combination of three above. The prediction performances of different time windows were also considered. Results showed that the uncertainty of user demand weakened the model performance and decreased the accuracy from about 90% to more than 60%. The accuracy of Bayesian classifier returned to 90% when the time window increased from 1 to 24. Quinones et al. [14] compared the performances of four machine learning models: kNN classifier, Bayesian classifier, artificial neural network (ANN) and SVM. Five cases of uncertainties were considered: leak size, measurement precision, pipe roughness, estimated user demand, and the combination of four above. It was found that almost 100% accuracies were achieved in the first three cases, while the uncertainty of user demand decreased the accuracies to more than 70%. The performances of models in the combination case were most weakened, and the accuracies dropped to about 60%. Moreover, the accuracies in the fifth case were improved by applying Bayes' rule in a time window. The accuracies of SVM and Bayesian classifier exceeded 90% when the time window was large enough. Lucin et al. [15] used a random forest model to predict leak locations. Three variables were randomly generated based on the Monk Carlo method: leak size, leak location, and user demand. The effect of data size was investigated. The prediction accuracy was improved as the data size increasing from 100,000 to 500,000. The comparison was carried out on the effects of uncertainties: leak size, user demand, sensor placement, sensor quantity and data feature. When multiple nodes with the top probabilities were considered, the prediction coverage of the real leak point was significantly improved to almost 100%, i.e., one of these predicted nodes is a real leak point.

In this study, a machine learning based method of block localization in WDNs is investigated. In a real network, the block data is insufficient to train machine learning model. Hence, numerous block scenarios were simulated by the hydraulic simulation considering three variables: user demand, block size and block location. The WDN model can be calibrated by the historical data from the real network. With a large range of simulation variables, the synthetic data was used to construct ANN models to predict the block locations. The influences of the variables were analyzed respectively.

2 Materials and Methods

2.1 Blockage Simulation

The simulation of the blockage has no common manner. In real WDNs, the pressure sensors are installed at the end of pipelines. The pressure distributions along the pipelines are not available. A study using the CFD method by Yang et al. [16] shows that the smaller the diameter and the longer the length of the narrow part, the larger the pressure drop. Thus, in this paper, blockage is simulated by reducing the pipeline diameter. The diameter coefficient C_d is proposed as a multiplier to reduce the pipeline diameter. The range of the diameter coefficient from 0.50 to 0.95 in a step of 0.01 was chosen arbitrarily. The range from 0.96 to 0.99 was not included considering subtle deviation and sensor error. The smaller diameter coefficient can cause significant resistance, which can be located by user reported water shortage. Thus, the diameter coefficients smaller than 0.50 were not considered.

2.2 Water Distribution Network Case

The case studied in this paper is the EPANET2 example Net3 network [17], a medium-sized WDN with 117 pipelines, 92 nodes, 3 tanks and 2 reservoirs (Fig. 1). The Net3 network case has 24 scenarios for every one hour in a day. Some simplifications were implemented to the network. The bypass pipelines of the pump 335 are canceled, and the lake source operates all day instead of opening in Hour 1-15. Thus, available nodes are 89 left, and pipelines are 115 left. Fundamental assumptions were considered that the hydraulic model is well calibrated, the blockage has been detected, and the blockage occurs in only one pipeline. The flow demands of end users can be satisfied and therefore fixed, and the pressure heads provided by water sources are also constant at fixed values.

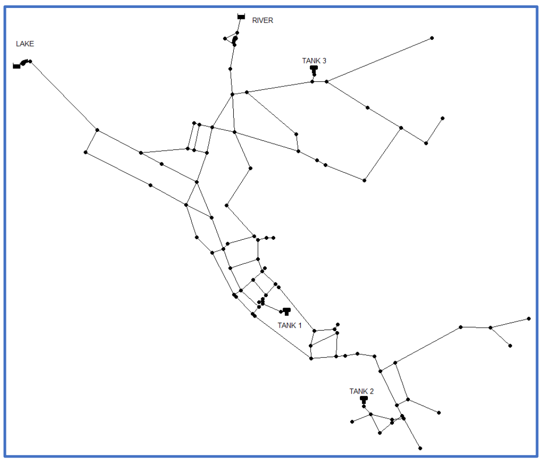

Fig. 1 The Net3 network layout

All the pipelines were considered being potentially blocked. In order to obtain the data for training the machine learning model, the block scenarios were conducted on a series of diameter coefficients and varying user demands. The network was modeled and simulated in CAENAE [18] in a time step of one hour. In this study, the simulation results of each hour were considered individually to investigate the model adaptability to various operating conditions. User demands change considerably, so the pressure change rates of the block scenarios relative to the normal scenarios were used. It is assumed that the flows and the pressures on the nodes can be measured in real time. The flow rates at the moment were considered as the parameters of the normal scenarios.

2.3 Artificial Neural Network

ANN has been commonly used in supervised prediction. ANN containing one hidden layer can achieve approximation to any non-linear function [19]. ANN gives out the probabilities of the labels, the final prediction is determined by the label of the maximum probability. Thus, ANN with one hidden layer was chosen to predict the block location. Adam solver was used for weight optimization and ReLU for activation function. The data features were built by the change rates of pressures at all nodes. Because the prediction of the blocked pipeline is a multi-classification problem, the accuracy is used as an efficient criterion to judge the model performance. The accuracy can be calculated by the proportion of true predictions. In this study, the dataset was divided by the hour and the step of the diameter coefficient, e.g., a train set of Hour 20 and Step 10 contains the data simulated with the user demands at 20 o'clock and the diameter coefficients of 0.50, 0.60, 0.70, 0.80, 0.90. And the test set of Hour 20 contains the data simulated with the diameter coefficients excluding 0.50, 0.60, 0.70, 0.80, 0.90.

The parameters of the ANN model (hyper-parameters) have a decisive influence on the prediction performance. Using the grid search method, the optimal hyper-parameters were determined with the dataset of Hour 0-5 and Step 10. With the constant learning rate and training epoch, the number of the hidden layer units and the L2 penalty were optimized for the ANN model. The hidden layer units of 80 and the L2 penalty of 0.6 were obtained. In Chapter 3, the ANN models were trained with the same hyper-parameters. The ANN models were realized with the Python library Pytorch 1.9.0 [20]. The flowchart of the proposed method is depicted in Fig. 2.

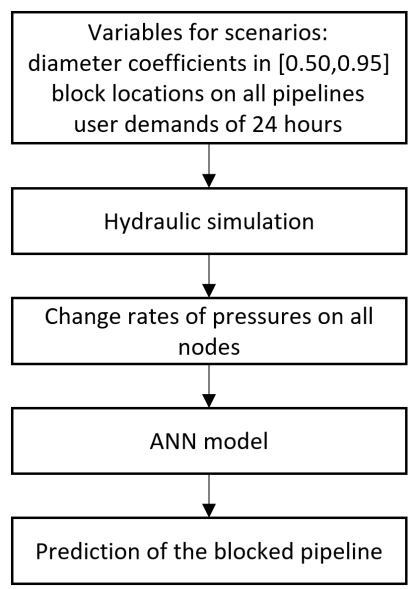

Fig. 2 Flowchart

3. RESULTS

3.1 User Demands Influence

The influence of user demands was investigated via various train sets. The dataset was divided into four parts by every six hours. The train sets are the first three parts, including Hour 0-17 scenarios, and the test sets are the fourth part, including Hour 18-23 scenarios. The model A/B/C was trained with the train set of Hour 0-5/0-11/0-17.

Model A achieved an overall accuracy of 72.9%, but only 55.9% for Hour 21 set. Model B showed an improvement, especially on Hour 21 set. The train set of model B included the scenarios similar to Hour 21 set and therefore the accuracy of Hour 21 set was improved significantly. Model C showed a slight improvement than model B, but the accuracies of Hour 22 set and Hour 23 set decreased. Scenarios under different user demands may have an opposite effect on the pressures and therefore decreased the accuracy.

Table 1. Influences of user demands.

Test set	Accuracy		
	Model A	Model B	Model C
Hour 18	71.9%	74.3%	78.7%
Hour 19	76.3%	80.2%	81.0%
Hour 20	73.2%	81.4%	81.6%
Hour 21	55.9%	74.8%	77.1%
Hour 22	83.8%	89.1%	88.4%
Hour 23	76.1%	76.9%	76.0%
Overall	72.9%	79.5%	80.5%

3.2 Diameter Coefficient Influence

To investigate the influence of the diameter coefficient, the dataset of Hour 0-5 was chosen and divided by the step of the diameter coefficient. The model D/E/F was trained with the train set of Step 10/5/2. The test set was the rest part of Hour 0-5 set. As the step was shortened, the data size of the train set increased.

Model D achieved an overall accuracy of 84.4% and model F reached the highest accuracy of 90.5%. With more training inputs, the model performance was improved. These models carried out the best results to the minimum diameter coefficient. As the diameter coefficient increased, the effect of the blockage became subtler and the accuracies decreased. Despite the maximum diameter coefficient, model F achieved an accuracy of 80.1%.

Table 2. Influences of diameter coefficient.

Test set	Accuracy		
	Model D	Model E	Model F
$0.50 < C_d < 0.60$	91.6%	93.2%	94.5%
$0.60 < C_d < 0.70$	90.1%	92.5%	94.6%
$0.70 < C_d < 0.80$	86.5%	90.2%	92.9%
$0.80 < C_d < 0.90$	80.1%	85.2%	88.6%
$0.90 < C_d \leq 0.95$	65.4%	76.2%	80.7%
Overall	84.4%	88.7%	90.5%

3.3 Block Location Influence

The model G was trained with the dataset of Hour 0-23 and the diameter coefficient of Step 5 to investigate the influence of the block location, achieving an accuracy of 86.6%. The accuracy for each pipeline was concluded in Table 3. For 65.2% of pipelines, the prediction results of model G were satisfactory, achieving an accuracy over 90%. The proportion of the pipelines with an accuracy over 50% reached 92.2%. The last 10 pipelines were shown in Table 4. Pipeline 293 and 319 were the most difficult to identify, which may need additional concerns.

Table 3. Proportions of the accuracies for each pipeline.

Accuracy range	Proportion
≥ 50%	92.2%
≥ 70%	82.6%
≥ 80%	74.8%
≥ 90%	65.2%

Table 4. Accuracies of the last 10 pipelines.

Pipeline	Accuracy	Pipeline	Accuracy
293	21.1%	211	45.3%
319	26.9%	281	46.6%
271	39.1%	135	47.3%
40	42.1%	50	48.2%
112	43.7%	323	50.4%

3.4 Model Application

The model H was trained with the dataset of all scenarios. Another test set considered 10% base demand variation was generated to test the model. The diameter coefficient of pipeline 271 was set to 0.80.

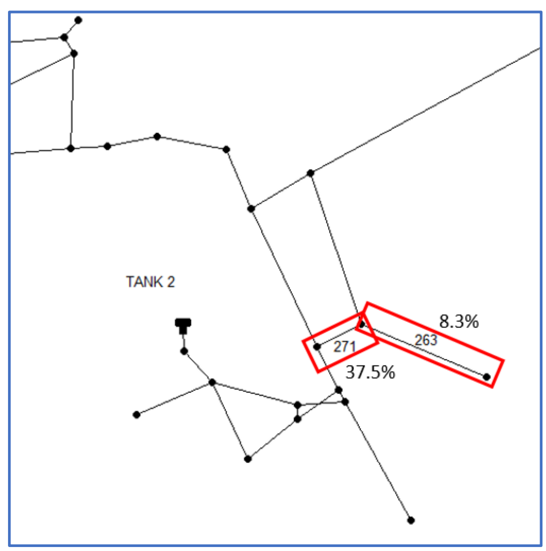

Fig. 3 Top 2 pipelines of predictions

Results of 24 times predictions showed that model H achieved an accuracy of 37.5% (decreased from 39.1%).

The second most predicted pipeline was 263 with a proportion of only 8.3% (Fig. 3). Hence, the final prediction can be determined by the majority voting. Although the accuracy was lower than 40%, model H was capable of locating the real blocked pipeline.

4. DISCUSSION

Based on the results obtained, it is evident that the proposed method can successfully locate a single blocked pipeline for a one-day operating scenario. Several models were constructed to compare the variables of the dataset. In conclusion, the more scenarios included in the train set, the better the performance of the model. However, with more scenarios, the computational cost gets higher, so a compromise should be considered between the model performance and the computational cost.

In this paper, the momentary user demands were considered as individual operating scenarios, and thus the data can be collected for multiple times in one day, with the final result obtained by the majority voting. Numerous flow and pressure sensors are required for this purpose. The influences of the number and the placement of the sensors should be investigated in further study. As described in Section 3.1, the user demands have a great impact on the performance of the model. The solution can be that building multiple models for varying user demand patterns.

There are still many open questions. Block detection in WDNs needs to be studied before block localization. The impact of different approaches to simulate blockage should be compared in further study, such as reducing the pipeline diameter and inserting a regulating valve in the pipeline, and the position of the valve may also affect the results. In the calibration of the hydraulic model, most pipelines may already be partially blocked, and thus the proposed method can only predict the further blocked pipelines on this basis. In this study, only a single blocked pipeline was considered. However, blockages caused by corrosion can occur in multiple pipelines at the same time, the prediction of multiple blocked pipelines should be studied in the future.

5. CONCLUSIONS

In this study, a machine learning based method was proposed to locate the blocked pipeline in the WDN. The dataset of the pressure change rates on the nodes was constructed by the hydraulic simulation, considering three variables: user demand, block size and block location. A series of machine learning models were built to investigate the influences of the dataset variables.

Results showed that the models were generally improved with a larger train set. However, the improvements are limited, which is helpful for the computational cost. 113 out of 115 pipelines obtained an accuracy higher than 30%. The model application indicated that the real block location can be determined by the majority voting based on a one-day measurement.

ACKNOWLEDGEMENT

This work was supported by National Key R&D Program of China (2021YFE0116200).

REFERENCE

[1] Duan H F, Lee P J, Ghidaoui M S, et al. Transient wave-blockage interaction and extended blockage detection in elastic water pipelines. Journal of Fluids and Structures 2014;46:2-16.

[2] Lile N L T, Jaafar M H M, Roslan M R, et al. Blockage Detection in Circular Pipe Using Vibration Analysis. International Journal on Advanced Science, Engineering and Information Technology 2012;2(3):252-255.

[3] Yang J, Feng Z, Wang X, et al. Research on Noise Reduction Method Based on CEEMD-WT-SVD and Its Application in Acoustic Signal of Pipeline Blockage. Journal of computers (China) 2019;30(2).

[4] Sattar A M, Chaudhry M H, Kassem A A. Partial Blockage Detection in Pipelines by Frequency Response Method. Journal of hydraulic engineering (New York, N.Y.) 2008;134(1):76-89.

[5] Duan H, Lee P J, Ghidaoui M S, et al. Extended Blockage Detection in Pipelines by Using the System Frequency Response Analysis. Journal of water resources planning and management 2012;138(1):55-62.

[6] Kim S. Multiple Discrete Blockage Detection Function for Single Pipelines. Proceedings 2018;2(11):582.

[7] Lee P J, Vítkovský J P, Lambert M F, et al. Discrete Blockage Detection in Pipelines Using the Frequency Response Diagram: Numerical Study. Journal of hydraulic engineering (New York, N.Y.) 2008;134(5):658-663.

[8] Massari C, Yeh T C J, Ferrante M, et al. A Stochastic Tool for Determining the Presence of Partial Blockages in Viscoelastic Pipelines: First Experimental Results. Procedia Engineering 2014;70:1112-1120.

[9] Massari C, Yeh T C J, Ferrante M, et al. A stochastic approach for extended partial blockage detection in viscoelastic pipelines: numerical and laboratory experiments. Journal of Water Supply: Research and Technology-Aqua 2015;64(5):583-595.

[10] Mashford J, De Silva D, Marney D, et al. An Approach to Leak Detection in Pipe Networks Using Analysis of Monitored Pressure Values by Support Vector Machine[C]. IEEE, 2009.

[11] Wachla D, Przystalka P, Moczulski W. A Method of Leakage Location in Water Distribution Networks using Artificial Neuro-Fuzzy System. IFAC-PapersOnLine 2015;48(21):1216-1223.
[12] Zhang Q, Wu Z Y, Zhao M, et al. Leakage zone identification in large-scale water distribution systems using multiclass support vector machines. Journal of Water Resources Planning and Management 2016;142(11):4016042.
[13] Soldevila A, Fernandez-Canti R M, Blesa J, et al. Leak localization in water distribution networks using Bayesian classifiers. Journal of Process Control 2017;55:1-9.
[14] Quiñones-Grueiro M, Bernal-De Lázaro J M, Verde C, et al. Comparison of Classifiers for Leak Location in Water Distribution Networks. IFAC-PapersOnLine 2018;51(24):407-413.
[15] Lučin I, Lučin B, Čarija Z, et al. Data-Driven Leak Localization in Urban Water Distribution Networks Using Big Data for Random Forest Classifier. Mathematics 2021;9(6):672.
[16] Yang L, Fu H, Liang H, et al. Detection of pipeline blockage using lab experiment and computational fluid dynamic simulation. Journal of Petroleum Science and Engineering 2019;183:106421.
[17] Zhao M, Zhang C, Liu H, et al. Optimal sensor placement for pipe burst detection in water distribution systems using cost–benefit analysis. Journal of Hydroinformatics 2020;22(3):606-618.
[18] Hai W, Zhengwei L. Complex water system analysis engine CAENAE-W: Principle, development and verification. Heating Ventilating & Air Conditioning 2021;51(01):16-22. (in Chinese)
[19] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Networks 1989;2(5):359-366.
[20] Paszke A, Gross S, Massa F, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. 2019.