WDSA CCWI 2022

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14766

ANALYSIS OF PDA-BASED WATER DISTRIBUTION SYSTEM SUSPENSION RISK USING STATISTICAL AND MACHINE LEARNING METHOD

Yoojin Oh[1], Haekeum Park[2], Jinseok Hyung[3], Taehyeon Kim[4], Kibum Kim[5] and Jayong Koo[6]

[1,2,3,4,6] University of Seoul, 219 room, 2 Engineering, Seoulsiripdae-ro 163, Dongdaemun-gu, Seoul, South Korea,
[5]Purdue University, HamptonHall of Civil Engineering 1237, United States of America

[1] *cordial5670@uos.ac.kr*, [2] *hkpark@uos.ac.kr*, [3] *gudwlstjr@uos.ac.kr*, [4] *kimth0712@uos.ac.kr*,
[5] *kim3854@purdue.edu*, 6 *jykoo@uos.ac.kr*

Abstract

Recently, there have been frequent cases of water shortages caused by failure to old water pipes. As water is the most basic resource in life and is an indispensable resource in various fields such as industry and agriculture, the scale of the failure is significant in the event of accidents in the water supply pipe network, and in order to minimize the damage of accidents, it is important to prevent accidents through timely maintenance. At this time, the risk map of the water shortage of the water pipeline needs to be prepared for efficient maintenance, and it needs to be managed first from the high-risk area.

To this end, water shortage risk analysis due to pipe failure was performed in this study. Risk analysis is one of the ways in which water pipes are evaluated and decisions on investment plans, such as replacement or repair, can be supported. The risk is generally calculated by multiplying PoF (Probability of failure) with the resulting direct and indirect effects of CoF (Consequence of failure). In this study, PoF was derived as the failure of an individual water pipe was set as the probability of failure caused by corrosion, and in order for it to be predicted, MLP (Multi-layer perceptron) and XGBoot were developed as a data-based machine learning model. In addition, it was analyzed by setting the amount of water (supply shortage) that CoF could not be supplied due to failure, considering that the failure to the water pipe was directly linked to water shortage. In order to analyze the supply shortage at this time, the mathematical analysis of PDA (Pressure driven analysis) was performed.

Finally, the developed methodology was applied to the cities of the Republic of Korea, and the risks were analyzed by calculating the PoF and CoF of individual water pipes, and the GIS technique was used to create the risk map.

The results of this study can be more accurate in predicting the condition of water pipes, which can be helpful when water utilities establish maintenance plans.

Keywords
Water Distribution System, Risk analysis, Risk map, Logistic regression, XGBoost regression, Pressure Driven Analysis(PDA).

1 INTRODUCTION

Recently, cases of water shortage caused by failure to old water pipes are occurring frequently. As water is the most basic resource in life and is an indispensable resource in various fields such as industry and agriculture, the scale of the failure is significant in the event of accidents in the water supply pipe network, and in order to minimize the damage of accidents, it is important to prevent accidents through timely maintenance.

In order to predict the condition of water pipes, the study of corrosion depth prediction models and failure rate prediction models using physical and statistical techniques has continued, and with the recent application of machine learning techniques, more accurate condition predictions have been attempted[1],[2],[3]. In previous studies, however, it has been difficult to develop physical models such as corrosion prediction models due to the limitation of data collection, and there is a limit that the correlation between corrosion and failure was not considered.

Risk, on the other hand, can be represented by the multiplication of PoF (Probability of failure) and CoF (Consequence of failure) of the direct and indirect effects resulting from it. So far, the risk analysis has only analyzed the possibility of failure using data on the history of failure to the pipe and considering that the accident in the water pipe is directly linked to the water shortage of the consumer, it is reasonable for the extent of the failure to be considered along with the probability of failure depending on the condition of the pipe. At this time, the scale of the failure can be calculated as supply shortage, which is the amount of water that is not supplied due to failure, and pressure driven analysis based on water pressure should be performed so that errors in interpretation such as the occurrence of negative pressure do not occur.

In this study, in order to evaluate risk based on the condition of the pipe, a corrosion depth prediction model according to the laying environment was developed to predict the depth of corrosion of individual water pipes, and the failure probability prediction model was developed by utilizing the results of the prediction and the history of failure in the past. In addition, the risk of individual pipelines was analyzed by calculating the supply shortage in case the pie is damaged through PDA mathematical analysis. Finally, the water shortage risk map of the water pipeline in the area subject to the study was drawn up.

2 METHODS

2.1 Study area

In this study, the research on the actual condition of the old water pipe was used to secure data on the influence factors of corrosion, and B city, where the history of failure to the water pipe is managed by GIS data, was selected as the target area of the study. Most of the water pipes laid in B city are DICP (94.1%), which were laid between 1968 and 2020. In addition, the date on the research on the actual condition is data surveyed between 2020 and 2021, which records the corrosion depth of the water pipeline and the environment in which it was laid, and a record of failure that occurred between 1999 and 2020.

2.2 Multi-Layer Perceptron

The artificial neural network model is a methodology that mimics the human brain, divided into single-layer perceptron and multi-layer perceptron according to the number of layers in which the neural network is constructed. Multi-layer perceptron model consists of input layer, hidden layer, and output layer, and between the nodes that make up each layer, weight is learned, and weighted sum is calculated. The output value of the weighted sum is calculated by activation function, and the output value is received as input value again, and the output value is calculated repeatedly through activation function.

$$z_j^k = w_{j,0}^{k-1} x_0^{k-1} + w_{j,1}^{k-1} x_2^{k-1} + \cdots w_{j,n}^{k-1} x_n^{k-1} \quad (1)$$

where, z_j^k : j-th node value of the k-th layer

$w_{j,n}^{k-1}$: Weight at which the n-th node of the (k-1)-layer and the j-th node of the k-layer are connected

x_n^{k-1} : n-th node value of the (k-1)-layer

2.3 XGBoost Algorithm

XGBoost stands for Extreme Gradient Boosting, which is one of the ensemble methods where Weak learner is combined sequentially to improve errors. Decision tree is the most widely used as the basic learner of XGBoost and is considered to show high performance as an overfitting-proof model due to the use of loss function and regularization term. When XGBoost's objective function is mathematically expressed, it is as the following (2) formula.

$$obj(\theta) = l(\theta) + \Omega(\theta) \tag{2}$$

where, $obj(\theta)$: Objective function

$l(\theta)$: loss function (Mean Square Error, MSE in case of regression, and loss of logistic in case of classification)

$\Omega(\theta)$: regularization term

2.4 Pressure Driven Analysis

In order to analyze the scale of the failure caused by the failure to the water pipe, a simulation through the interpretation of pipe network should be carried out, and it is necessary to conduct a PDA (Pressure driven analysis), not a DDA (Demand driven analysis), in which unrealistic results such as the occurrence of negative pressure can be produced under certain conditions. In particular, the consumer's supply capacity is determined by HOR (Head-Outflow Relationship), and the water supply is impossible if the consumer's water pressure becomes less than the minimum water pressure, and if the head of the consumer's pressure falls between the minimum and marginal water pressure, the supply is supplied according to the HOR, providing only a fraction of the required demand supply.

$$\begin{cases} q_j^{avi} = q_j^{req}, & if\ H_j^{avi} \geq H_j^{des} \\ 0 < q_j^{avi} = q_j^{req}\left(\dfrac{H_j^{avi} - H_j^{min}}{H_j^{des} - H_j^{min}}\right) < q_j^{req}, & if\ H_j^{min} < H_j^{avi} < H_j^{des} \\ q_j^{avi} = 0, & if\ H_j^{avi} \leq H_j^{min} \end{cases} \tag{3}$$

3 RESULTS

3.1 Probability of Failure

The corrosion depth prediction model was developed to predict the probability of failure caused by corrosion, and the result was used as an input variable for the failure probability prediction model. Parameters used for the development of the corrosion depth prediction model and the failure probability prediction model are as follows in Table 1. The input variables of each model were selected for items of high correlation and items mainly used in prior studies as a result of the analysis of the correlation with the output variables. Hyperparameters were set through Bayesian Optimization techniques.

Table 1. Parameters of corrosion depth prediction model and failure probability prediction model

Classification	Input		Output
Corrosion depth prediction model	Pipe data	Pipe age, pipe material, pipe diameter	Corrosion depth
	Laying environment	Laying depth, sulphide concentration, soil pH, oxidation-reduction potential, moisture content	
	Operating conditions	Water quality corrosiveness (LI), maximum water pressure	
Failure probability prediction model	Pipe data	Pipe age, pipe material, pipe diameter, joint type, pipe thickness (predicted value)	Failure probability
	Laying environment	Laying depth, road shape	
	Operating conditions	Average flow rate, maximum water pressure	
	Failure history	The number of failure in the last 5 years, the number of failure in the last 10 years	

Training set and Test set are divided into 8:2, and the result of the development of the corrosion depth prediction model and the failure probability prediction model is the same as Table 2 and Table 3. In the evaluation of the model, the coefficient of determination (R^2), and RMSE were used.

Table 2. Results of the development of corrosion depth prediction model and failure probability prediction model (Regression)

Classification	Method	Training set		Test set	
		R^2	RMSE	R^2	RMSE
Corrosion depth prediction model	MLP	0.88	1.6	0.68	1.92
	XGBoost	0.90	1.45	0.84	1.41
Failure probability prediction model	MLP	0.97	2.23	0.42	8.64
	XGBoost	0.89	0.05	0.69	0.08

In the case of the corrosion depth prediction model, the Test set's R^2 was 0.84 and the RMSE was 1.41, which was confirmed to predict the depth of corrosion in the water pipe more accurately than the MLP model. In the failure probability prediction model, the XGBoost model more accurately predicted than the MLP model, with Test set's R^2 was 0.65 and RMSE was 0.19 for the probability of failure to the water pipe. The developed model was applied to B city to derive the probability of failure to individual water pipes.

3.2 Consequence of Failure

In this study, EPANET 2.2 toolkit was used to calculate the supply shortage caused by failure to water pipes, which simulated the failure of individual pipelines. First of all, the supply shortage was calculated in consideration of the time of recovery from the time the failure occurred based on the maximum supply of water in normal condition where no failure occurred. At this time, the results of the survey of the time of recovery by pipe diameter in the past were utilized as the time of recovery. The analysis showed that the supply shortage in case of failure was greater in the main drain than in the pipe-end area, and that the blocking made it possible to supply water despite failure to the pipe, leaving some areas where water shortage does not occur.

3.3 Risk analysis and risk map

Finally, the risk of individual pipelines multiplied by the impact (supply shortages) resulting from failure probability and failure was calculated. The result of visualizing the zone-specific risk of B city is as follows in Figure 1. The characteristics of pipe laid in the highest and lowest risk areas have been analyzed, and it has been confirmed that pipe network's blocking is constructed even if the probability of failure to pipe is high, resulting in a lower score if water shortage does not occur.

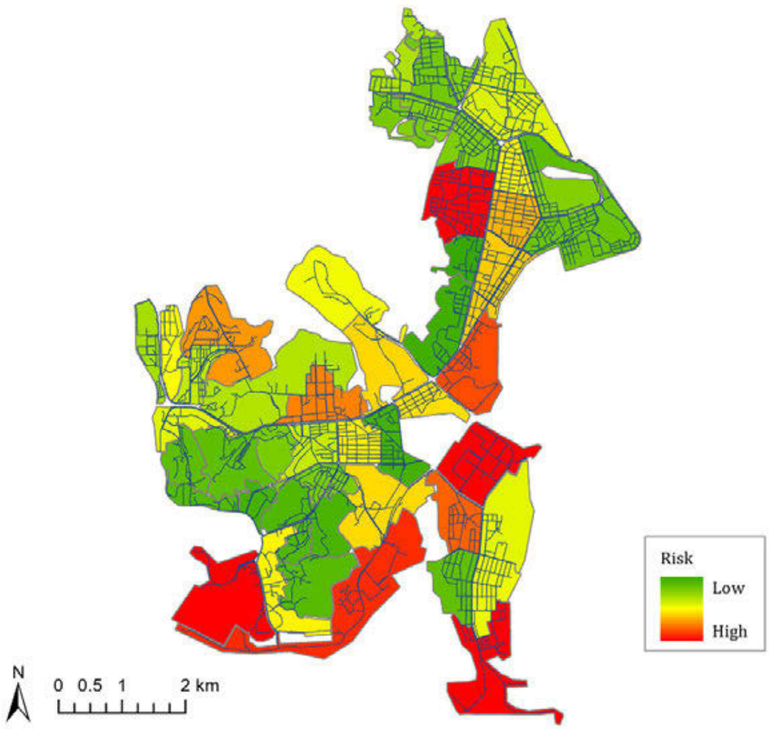

Figure 1. Risk map of B city

4 CONCLUSIONS

In this study, corrosion depth prediction model and failure probability prediction model were developed as data from research on the actual condition in water pipes and failure history data were used to analyze the risk of water shortage caused by corrosion of water pipes. As a result of the development of the model, the performance of the model using XGBoost was shown to be superior.

In addition, the PDA mathematical analysis technique was used to quantify the impact of failure on water pipes, resulting in a shortage of supply in the event of failure to individual pipelines. In addition, the PDA mathematical analysis technique was used to quantify the impact of the failure of the water pipeline, calculating the supply shortage in the event of failure to individual pipelines. Finally, as the probability of failure of individual pipelines was multiplied by the impact of failure (supply shortage), the risk of water shortage in the individual pipeline was calculated, and the zone-specific risk was visualized, and risk map was expressed.

This study can be helpful when water utilities identify the status of pipes and determine areas where maintenance is needed first.

5 ACKNOWLEDGEMENTS

The Korea Ministry of Environment supported this work titled as "Project for developing innovative drinking water and wastewater technologies (2020002700016)".

6 REFERENCES

[1] Uri Shamir and Charles D.D. Howard, "An Analytic Approach to Scheduling Pipe Replacement," Journal AWWA, vol. 71, no. 5, 1979, pp. 248-258.

[2] Robert M. Clark, "Water Distribution Systems : A Spatial and Cost Evaluation," Journal of Water Resources Planning and Management, vol. 108, no. 3, 1982, pp. 243-256.

[3] Xudong Fan, Xiaowei Wang, Xijin Zhang, P.E.F. ASCE Xiong (Bill) Yu, "Machine learning based water pipe failure prediction: The effects of engineering, geology, climate and socio-economic factors," Reliability Engineering and System Safety, vol. 219, no. 108185, 2022, pp. 1-15.