engineering proceedings

Proceeding Paper

Optimal Sensor Placement in Water Distribution Networks Using Dynamic Prediction Graph Neural Networks †

Aly K. Salem and Ahmed A. Abokifa *

Department of Civil, Materials, and Environmental Engineering, The University of Illinois Chicago, Chicago, IL 60607, USA; asalem22@uic.edu

* Correspondence: abokifa@uic.edu; Tel.: +1-312-413-4636

† Presented at the 3rd International Joint Conference on Water Distribution Systems Analysis & Computing and Control for the Water Industry (WDSA/CCWI 2024), Ferrara, Italy, 1–4 July 2024.

Abstract: Sensors are a key component of water distribution networks due to their role in monitoring system variables. Specifically, water quality (WQ) sensors are utilized to measure chlorine concentrations in order to maintain water quality standards. However, the prohibitive costs of deploying these sensors constrain their ubiquitous use. As a result, WQ sensors are typically placed in a subset of junctions that are selected via an optimization process. This study presents a framework for optimizing WQ sensor placement to maximize chlorine concentration state estimation, that is, the inference of water quality parameters at unmonitored junctions based on the measurements from monitored junctions. This is performed by integrating a Dynamic Prediction Graph Neural Network (DP-GNN) model with a Genetic Algorithm (GA). The DP-GNN model is trained to predict chlorine concentrations at all junctions based on the measurements from sensors with different placements, whereas the GA uses these predictions to find the optimal sensor placement. The framework performance was tested by applying it to the C-town benchmark network, considering different numbers of sensors. The results demonstrated the impact of different sensor placements on the prediction accuracy of the DP-GNN model. Additionally, the results showed the framework's ability to find the sensor placement that maximizes the chlorine concentration state estimation performance.

Keywords: sensor placement; optimization; water systems; water quality; graph neural networks

Citation: Salem, A.K.; Abokifa, A.A. Optimal Sensor Placement in Water Distribution Networks Using Dynamic Prediction Graph Neural Networks. *Eng. Proc.* **2024**, *69*, 171. https://doi.org/10.3390/engproc2024069171

Academic Editors: Stefano Alvisi, Marco Franchini, Valentina Marsili and Filippo Mazzoni

Published: 25 September 2024

1. Introduction

Sensors are essential for the effective operation and management of water distribution networks (WDNs). Sensors serve in monitoring flows, pressures, and detecting leaks [1], as well as regulating network components such as pumps and valves. Recently, the reliance on water quality (WQ) sensors has increased due to escalating concerns over public health risks associated with accidental or intentional contamination within WDNs [2], in addition to the gradual deterioration of water quality within these networks [3]. These sensors serve in contamination detection and chlorine concentration monitoring. However, deploying WQ sensors in WDNs poses a challenge due to their higher cost compared to flow and pressure sensors [1]. To face this challenge, water quality sensors are strategically placed at selected junctions to reduce costs without compromising the performance of the sensor network [3]. This placement process is commonly approached as an optimization problem, formally known as the sensor placement optimization (SPO) problem.

Several studies targeted the SPO problem by developing simulation–optimization approaches to maximize contamination detection [2]. This approach utilizes physics-based models, such as EPANET, for water and contamination transport simulations alongside optimization algorithms for objective function optimization. The proposed methodologies vary based on the number and nature of the objective functions, categorized as single or multi-objective approaches. A comprehensive review of SPO methodologies for contamination detection was conducted by Hart and Murray [4]. In contrast, fewer studies have focused on SPO for chlorine concentration estimation. In recent work, Łangowski and Brdys [3] and Taha et al. [5] proposed methods to optimize WQ sensor placement to minimize chlorine concentration prediction errors through Kalman filtering and a multi-objective optimization approach, respectively.

Recently, Graph Neural Networks (GNNs) were introduced as an advancement over traditional Artificial Neural Networks in handling complex graph-structured data. GNNs excel in learning the underlying relationships between nodes and edges of a graph, making them suitable for WDN applications. As a result, GNNs have been utilized in contamination source identification and pressure and water loss estimation [6–8]. However, the application of GNNs in WQ sensor placement optimization has not been explored yet. In this paper, we aim to fill this gap by investigating the implementation of GNNs in solving the SPO problem. This is achieved by developing an SPO framework integrating Dynamic Prediction Graph Neural Networks (DP-GNN) [9] with a Genetic Algorithm (GA). This framework aims to identify the optimal sensor placement that maximizes the accuracy of chlorine concentration estimation based on limited sensor data.

2. Materials and Methods

The sensor placement optimization (SPO) framework integrates two models: (i) the simulation model and (ii) the optimization model. The simulation model features a Dynamic Prediction Graph Neural Network (DP-GNN) model, while the optimization model is represented by the Genetic Algorithm (GA). The DP-GNN is tailored to predict chlorine concentrations (CCs) based on the data provided by any sensor configuration (i.e., sensor design). It is constructed by stacking multiple Topology Adaptive Graph Convolution Network (TAGCN) layers [10] alongside dense layers. A brief explanation of the DP-GNN model is given in this paper, whereas a detailed explanation of the model assumptions, structure, and hyperparameters can be found in [9].

2.1. Simulation Model Training

To train the DP-GNN, a large dataset of events encompassing varying junction demands, sensor designs, and chlorine injection rates was randomly generated. These events were then simulated by the Python interface of EPANET (WNTR) to extract CC data from all junctions. Since the DP-GNN model is supposed to learn the relationship between the sensor and non-sensor junctions, the latter were introduced by masking the CC data for selected junctions. This masking process involves randomly selecting sensor numbers and locations, thereby masking the CC of the remaining junctions.

The DP-GNN assumes that flow sensors exist at all junctions; hence, the demands are known, resulting in an input vector of $[Q_n, C_n,\ J_n]$ for each junction within an event (i.e., graph). In this vector, Q denotes the demand, C represents the CC value or 0 if unknown, and J indicates the junction type, a binary value which is set to 1 for monitored junctions and 0 otherwise. In addition to the junction input vector, the edge adjacency matrix (A) that represents the node-to-node connection through graph edges is defined in the DP-GNN model. The DP-GNN model was trained to reduce the normalized root mean square error $(nRMSE)$ shown in Equation (1). The assessment of the DP-GNN model performance for individual junctions in different events (i.e., graphs) was performed using the mean absolute percentage error ($MAPE$) described by Equation (1).

$$nRMSE = \frac{\left[\sum_{g=1}^{G}\sum_{n=1}^{N}\left(C_n - \hat{C}_n\right)^2/N\right]^{1/2}}{\sum_{g=1}^{G}\left(\sum_{n=1}^{N}C_n\right)/N} \quad (1)$$

where C_n and $\hat{C}_n$ are the actual and predicted chlorine concentrations, and G and N are the total number of graphs and junctions.

2.2. Optimization Model

The GA was utilized in this paper due to its wide use in SPO studies [1–3]. The GA follows the natural selection mechanisms to evolve from suboptimal solutions to optimal ones. In the context of this paper, the GA starts by providing several sensor placements (i.e., sensor designs) to the trained DP-GNN. Subsequently, the DP-GNN utilizes this information to define the sensor numbers and locations. Then, it runs to predict the CCs for non-sensor junctions, calculate the $nRMSE$, and report it back to the GA. The GA minimizes the $nRMSE$ through iterative refinement to identify the optimal sensor design. The GA parameters used in this study are shown in Table 1.

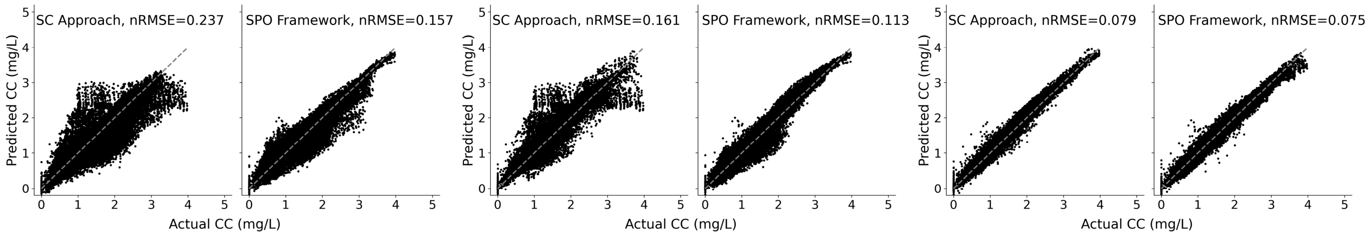
Table 1. Genetic Algorithm optimization parameters used in the paper.

Optimization Parameter	Value	Optimization Parameter	Value
Population size	5000	Crossover probability	0.8
Number of generations	100	Mutation probability	0.2
Parents percentage	1%		

3. Results and Discussion

The proposed SPO framework was applied to the C-town benchmark network detailed in [9] to test its performance. Three different scenarios were examined to represent the diverse sensor placement criteria commonly encountered in real-world situations. In scenario 1 (S1), 3 sensors were assumed to be placed in the network, whereas 5 and 10 sensors were assumed in S2 and S3, respectively. The SPO framework results were assessed against the sensor placements generated by the Spectral Clustering (SC) approach. In these scenarios, 1000 different events encompassing a range of demands and chlorine injection rates were considered.

Figure 1 illustrates the one-to-one plot depicting the results for each scenario, contrasting the sensor placements derived from the SC approach against those generated by the SPO framework. The x-axis represents the actual chlorine concentrations (CCs), while the y-axis represents the corresponding CC predictions made by the DP-GNN model. Additionally, the $nRMSE$ is indicated in each subplot, providing a quantitative measure of the prediction accuracy.



(**a**) (**b**) (**b**)

Figure 1. The one-to-one plot of the SC approach compared to the SPO framework for (**a**) S1, (**b**) S2, and (**c**) S3.

By examining the left and right panels of each subplot, it becomes evident that the sensor design proposed by the SPO framework exhibits lower $nRMSE$ compared to those derived from the SC approach. This trend persists across all three scenarios. Furthermore, Figure 1 shows that the increased number of sensors (e.g., from S1 to S2) resulted in more accurate DP-GNN model predictions. Interestingly, the optimal sensor design of S1 outperformed the sensor design produced by the SC approach in S2, although fewer sensors were used. This highlights the superiority of the SPO framework over the clustering approach. Nevertheless, the disparity in $nRMSE$ between the two approaches diminishes with the addition of more sensors. This phenomenon can be attributed to the diminishing benefit of optimization when the objective function is insensitive to the design variables.

Author Contributions: Conceptualization, A.K.S.; methodology, A.K.S.; software, A.K.S.; validation, A.K.S.; formal analysis, A.K.S.; investigation, A.K.S.; resources, A.K.S.; data curation, A.K.S.; writing—original draft preparation, A.K.S.; writing—review and editing, A.A.A.; visualization, A.K.S.; supervision, A.A.A.; project administration, A.A.A.; funding acquisition, A.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Science Foundation (NSF) under Grant No. 2015603.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this paper are available upon reasonable request from the corresponding author.

Acknowledgments: Support from the NSF is gratefully acknowledged.

Conflicts of Interest: The authors declare no competing interests.

References

1. Casillas, M.V.; Puig, V.; Garza-Castañón, L.E.; Rosich, A. Optimal sensor placement for leak location in water distribution networks using genetic algorithms. *Sensors* **2013**, *13*, 14984–15005. [CrossRef] [PubMed]
2. Aral, M.M.; Guan, J.; Maslia, M.L. Optimal Design of Sensor Placement in Water Distribution Networks. *J. Water Resour. Plan. Manag.* **2010**, *136*, 5–18. [CrossRef]
3. Łangowski, R.; Brdys, M.A. An optimised placement of the hard quality sensors for a robust monitoring of the chlorine concentration in drinking water distribution systems. *J. Process Control* **2018**, *68*, 52–63. [CrossRef]
4. Hart, W.E.; Murray, R. Review of Sensor Placement Strategies for Contamination Warning Systems in Drinking Water Distribution Systems. *J. Water Resour. Plan. Manag.* **2010**, *136*, 611–619. [CrossRef]
5. Taha, A.F.; Wang, S.; Guo, Y.; Summers, T.H.; Gatsis, N.; Giacomoni, M.H.; Abokifa, A.A. Revisiting the Water Quality Sensor Placement Problem: Optimizing Network Observability and State Estimation Metrics. *J. Water Resour. Plan. Manag.* **2021**, *147*, 04021040. [CrossRef]
6. Li, Z.; Liu, H.; Zhang, C.; Fu, G. Gated graph neural networks for identifying contamination sources in water distribution systems. *J. Environ. Manag.* **2024**, *351*, 119806. [CrossRef] [PubMed]
7. Ashraf, I.; Hermes, L.; Artelt, A.; Hammer, B. Spatial Graph Convolution Neural Networks for Water Distribution Systems. *Lect. Notes Comput. Sci.* **2023**, *13876*, 29–41.
8. Fu, M.; Zhang, Q.; Rong, K.; Mundher, Z.; Zheng, L. Integrated dynamic multi-threshold pattern recognition with graph attention long short-term neural memory network for water distribution network losses prediction: An automated expert system. *Eng. Appl. Artif. Intell.* **2024**, *127*, 107277. [CrossRef]
9. Salem, A.K.; Taha, A.F.; Abokifa, A.A. Graph Neural Networks-Based Dynamic Water Quality State Estimation in Water Distribution Networks. *Research Square* **2024**. [CrossRef]
10. Du, J.; Zhang, S.; Wu, G.; Moura, J.M.F.; Kar, S. Topology Adaptive Graph Convolutional Networks. *arXiv* **2017**, arXiv:1710.10370.