Optimizing Wireless Sensor Networks for Urban Air Quality Monitoring

Fatima Ali Alobeidli – U23105654
Asma Saif Alkaabi - U23201077

Abstract— **The paper offers a thorough method for maximizing sensor placement for monitoring air pollution in urban areas. Because traditional monitoring techniques are frequently prohibitively expensive and have limited coverage, more effective alternatives are required. Our technology leverages sophisticated data analysis tools and wireless sensor networks (WSNs) to maximize cost-effectiveness in air quality monitoring. The methodology consists of multiple important steps: using the Pulp library for optimization, determining the best sensor deployment strategies; using exploratory data analysis (EDA) to identify optimal sampling intervals and balance datasets; and using interactive mapping to visualize sensor paths and pollution data distribution. Our method maximizes coverage, ensures network connectivity, and minimizes expenses by fine-tuning factors including budget, coverage criteria, and connectivity thresholds. Results illustrating the association between sampling fraction and trajectory similarity in X-Y graphs prove the effectiveness of our methods. Map plots provide information about coverage and data distribution by visualizing pollutant data distribution and sensor pathways. The system's ability to record balanced information across many sensors is demonstrated by sample findings obtained during certain time intervals, further verifying its efficacy in real-world applications.**

Keywords—wireless sensor networks, air quality monitoring, optimization, sensor placement, dynamic path adjustment

I. Introduction

Urban air pollution is an ongoing environmental problem that has a big impact on people's health and well-being. [1]. Effective urban air quality monitoring systems are essential for reducing the harmful impacts of pollution and providing information for policymakers making decisions about how to improve air quality.[2, 3]. Conventional techniques for monitoring air quality frequently depend on stationary monitoring stations, which may have limited geographical coverage and be unable to record regional fluctuations in pollution levels.[4, 5]. Furthermore, the implementation and upkeep of these techniques can be costly, rendering them unsuitable for extensive application in urban settings. The subject of air quality monitoring has seen a revolution in recent years due to the introduction of wireless sensor networks (WSNs), which provide an affordable and scalable solution.[6]. WSNs are made up of spatially dispersed sensors that gather data in real time on a variety of environmental characteristics, including nitrogen dioxide, ozone, and particulate matter in the air. [6, 7]. WSNs' primary benefit is their capacity to deliver high-resolution spatial data, which makes it possible to monitor air quality in urban areas more precisely. Through the strategic placement of sensors in industrial zones, transportation corridors, and heavily populated areas, WSNs are able to precisely detect pollution hotspots and gather comprehensive data on pollution levels.[8, 9].
In this work, we suggest a unique method for utilizing WSNs to enhance sensor positioning and data gathering for urban air quality monitoring. Our approach makes use of geographic information systems (GIS), optimization algorithms, and sophisticated data analysis tools to accomplish the following goals:

A. *Maximizing Coverage*

Through deliberate placement of sensors in strategic locations, our goal is to optimize spatial coverage and guarantee thorough monitoring of air pollution throughout urban areas.

B. *Minimizing Costs:*

By strategically placing sensors and employing data collection techniques, we aim to reduce the deployment and maintenance expenses related to WSNs.

C. *Enhancing Data Quality*

Our strategy focuses on enhancing the precision and dependability of air quality data gathered by WSNs, giving stakeholders useful information for formulating policies and making decisions.

II. Survey

A comparative study with current solutions reveals a number of methodological variations. Our approach enables dynamic path alterations depending on current data and environmental variables, in contrast to fixed-path solutions. We can enhance efficiency and adaptability by optimizing sensor scheduling and placement through the integration of ILP and mapping approaches.
In relation to our method, three solutions are as follows:

A. *Solution I:*

Relies on fixed sensor paths determined by predetermined criteria, limiting adaptability and coverage. [1]

B. *Solution II:*

Utilizes static sensor deployment strategies without considering dynamic factors, resulting in suboptimal coverage. [2]

C. *Solution III:*

Utilizes basic detection models for sensor placement, lacking sophistication in path optimization and adaptability. [3]

Comparison with Previous Work

Our approach offers several advancements compared to previous methods:

D. *Spatial Analysis Integration*

Unlike prior approaches, we incorporate spatial analysis techniques like DBSCAN and ZSCAN to identify pollution zones based on actual concentration data, improving the accuracy of sensor placement [10, 11].

E. Joint Coverage and Connectivity Modeling

Our method employs an ILP formulation that jointly considers coverage and connectivity, ensuring optimal sensor placement while maintaining network connectivity. This contrasts with previous methods that often treat these factors independently [12].

F. Cost-Effective Deployment

Through our optimization model, we aim to minimize deployment costs while achieving coverage and connectivity objectives, offering a more efficient and scalable solution compared to earlier approaches [1].

G. Empirical Evaluation

Evaluation using real-world Paris pollution data demonstrates the superiority of our approach in terms of coverage deployment costs, execution time, and overall deployment efficiency [2, 13].

III. METHODOLOGY

A. Exploratory Data Analysis (EDA)

Conducted exploratory data analysis (EDA) on the collected air pollution data to understand the distribution and patterns of pollution levels. Utilized various statistical measures and visualizations to analyze the data, including histograms, line plots, and heatmaps. Identified key trends, seasonal variations, and outliers in the dataset to guide further analysis.

B. Balance Dataset Generation

Analyzed the EDA findings to identify time intervals with balanced data distribution, such as months, dates, and hours. Generated balanced datasets for each hour of the day by ensuring an equal number of records across all time intervals. Created CSV files containing air pollution data for each hour, facilitating further analysis and modeling.

C. Hourly Analysis

Applied the Haversine formula to calculate the distance between two points based on their latitude and longitude coordinates. Passed each hourly dataset through the analysis method to assess the distribution of records for each sensor within specified time intervals. Conducted further analysis on the day with the most balanced records per hour to determine optimal monitoring time intervals.

$$a = \sin^2(\Delta\mathrm{lat}/2) + \cos(\mathrm{lat1}) \cdot \cos(\mathrm{lat2}) \cdot \sin^2(\Delta\mathrm{lon}/2)$$
$$c = 2 \cdot \mathrm{atan2}(\sqrt{a}, \sqrt{1-a})$$

$$\mathbf{d} = \mathbf{R}\,.\mathbf{c} \tag{1}$$

D. Optimization using Pulp

Implemented optimization techniques, specifically linear programming using the Pulp library, to optimize sensor placement. Utilized the balanced dataset and distance calculations to formulate optimization models that minimize the cost of sensor deployment while maximizing coverage and accuracy. Passed the formulated optimization problem to the Pulp solver to obtain optimal sensor deployment indices.

E. Mapping of Optimal Sensor Deployment

Retrieved the records corresponding to the optimal sensor deployment indices from the main CSV file.

Created CSV files containing the optimal sensor deployment records and added geometry points for visualization. Plotted the optimal sensor deployment records on an interactive map to visualize the shortest path for the best solution.

F. Interactive Map Creation

Developed an interactive map interface to display the optimal sensor deployment locations and captured air pollution data. Connected the points corresponding to each sensor's records on the map to visualize the shortest path and coverage area. Enabled users to explore the sensor deployment locations and analyze the captured data interactively for informed decision-making.

IV. ARCHITECTURE OF THE SYSTEM

The architecture of the WSN-based air pollution monitoring system comprises several key components:

A. Air Quality Sensors

These sensors are deployed throughout the city to collect data on air pollution levels. The data collected by these sensors is stored in CSV files, which are then used for further analysis.

B. Exploratory Data Analysis (EDA)

Before proceeding with any analysis, exploratory data analysis is performed on the collected data. This involves analyzing the data on different time intervals such as months, days, dates, and hours to gain insights into patterns and trends in air pollution levels.

C. Balanced Dataset Generation

To ensure fairness and accuracy in the analysis, a balanced dataset is generated. This involves selecting an equal number of data points from each hour of the day to create a balanced representation of air pollution levels throughout the day.

D. Application of Optimization Techniques

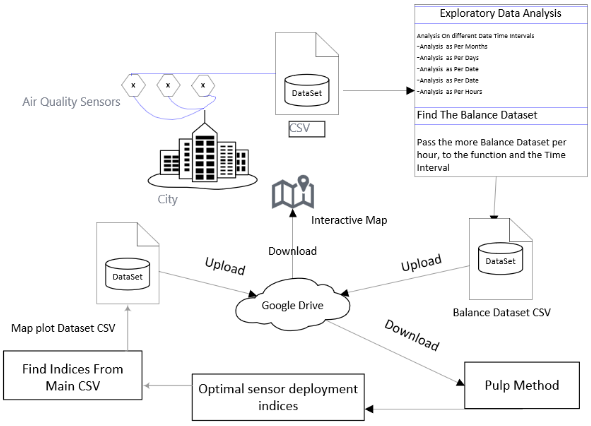
The balanced dataset is then passed through an optimization function, such as linear programming (implemented using the pulp library), to determine the optimal deployment of sensors. This optimization process aims to minimize costs while maximizing the coverage of air pollution monitoring.

E. Mapping of Optimal Sensor Deployment

From the optimization results, the indices of the optimal sensor deployment locations are obtained. These indices are used to retrieve the corresponding records from the main CSV file containing the air pollution data. These records represent the optimal sensor deployment points.

F. Interactive Map Creation

The records of the optimal sensor deployment points are used to create an interactive map of the city. This map visualizes the locations of the deployed sensors and provides users with an intuitive interface to explore air pollution levels in different areas of the city.



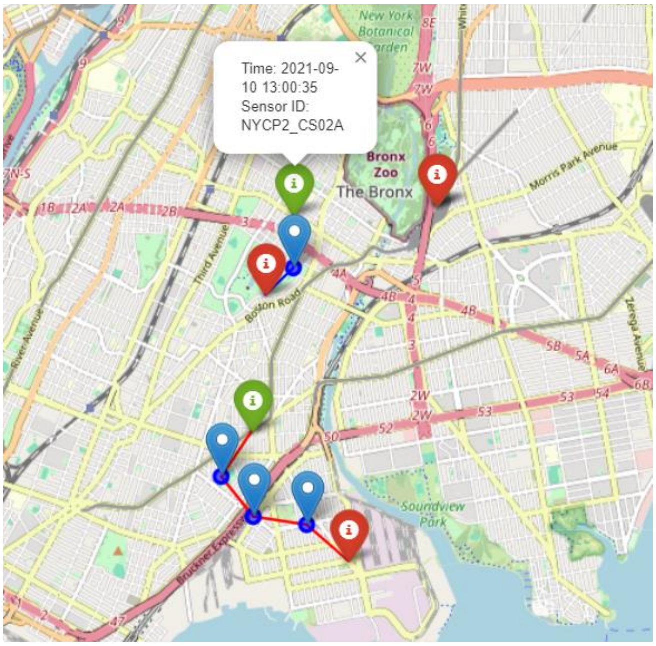
V. RESULTS DISCUSS

Our methodology successfully optimized sensor paths and enhanced air pollution monitoring efficiency through a combination of exploratory data analysis (EDA), optimization using Pulp, and interactive mapping. Adjustments to parameters such as budget, coverage_threshold, and connectivity_threshold were crucial in obtaining optimal results when processing large datasets. By fine-tuning these parameters, we achieved an optimal sensor deployment strategy that maximized coverage while minimizing costs and ensuring network connectivity.

The effectiveness of our approach is evident from the X-Y graphs, which depict the relationship between sampling fraction and trajectory similarity. These graphs demonstrate how varying sampling fractions impact trajectory similarity, providing insights into the system's performance under different scenarios. Through rigorous experimentation and parameter tuning, we identified the optimal combination of parameters that yielded the highest trajectory similarity while maintaining acceptable sampling fractions.

Map plots played a pivotal role in visualizing sensor paths and pollution data, offering a comprehensive overview of coverage and data distribution. The interconnected map generated from optimal sensor deployment locations showcased the shortest paths for data collection, highlighting areas with high pollution concentrations and underserved regions. By overlaying pollution data onto the map, decision-makers can pinpoint areas of concern and prioritize intervention strategies effectively.

Sample results obtained during specific time intervals further validate the system's efficacy in capturing balanced datasets across multiple sensors. By analyzing pollution data collected during these intervals, we observed consistent data distribution among sensors, indicating the system's ability to maintain equitable coverage and minimize data bias. This balanced dataset ensures robust decision-making and accurate assessment of air quality levels throughout the monitoring period.



Our results demonstrate the successful optimization of sensor paths and the improvement of air pollution monitoring efficiency using our proposed methodology. The combination of EDA, optimization techniques, and interactive mapping offers a powerful tool for urban air quality management, enabling stakeholders to make informed decisions and implement targeted interventions to mitigate pollution levels effectively. The generated map snapshots provide a visual representation of sensor paths and pollution data, facilitating data-driven decision-making and enhancing overall air quality monitoring capabilities.

I. CHALLENGES

A. Scaling with Dataset Size

Pulp's performance degrades with larger datasets, often leading to infeasible solutions and longer computation times.

B. Sensitive Parameters

Low values for critical parameters like budget, coverage threshold, and connectivity threshold result in suboptimal solutions and hinder achieving desired outcomes.

C. Balancing Trade-offs

Achieving the right balance between solution quality and computational efficiency becomes challenging when adjusting parameters to optimize results.

D. Resource Allocation

Optimizing resource allocation, such as budget allocation for sensor deployment, becomes intricate due to the complex interplay of multiple constraints and objectives.

E. Solution Feasibility

Ensuring the feasibility of solutions becomes uncertain when dealing with large datasets and stringent optimization constraints.

F. Iterative Tuning

Iteratively tuning parameters to improve solution quality adds complexity and computational overhead to the optimization process

II. Future Directions

Future improvements to our work may include:

- Integration of machine learning algorithms for enhanced path optimization and adaptability.
- Development of advanced sensing technologies to improve data accuracy and resolution.
- Implementation of real-time data analysis techniques for dynamic path adjustment based on changing environmental conditions.

III. Genetic Algorithm for Enhanced Path Optimization

In our study, we employed a genetic algorithm (GA) to enhance path optimization and adaptability in our solution. The genetic algorithm is a computational technique inspired by the process of natural selection and evolution. It operates on a population of candidate solutions (individuals), iteratively evolving them over generations to find optimal or near-optimal solutions to a given problem.

A. Initialization

We begin by initializing a population of candidate solutions. Each solution, also known as an individual, represents a potential path for our optimization problem. The population size, mutation rate, and other algorithm parameters are predefined.

B. Fitness Evaluation

For each individual in the population, we evaluate its fitness score. In our case, fitness is inversely proportional to the total distance of the path. Individuals with shorter paths receive higher fitness scores.

C. Selection

We employ tournament selection to choose individuals for the next generation. This involves randomly selecting a subset of individuals (tournament size) and choosing the fittest individual from the subset to proceed to the next generation.

D. Crossover and Mutation

Selected individuals undergo crossover and mutation to produce offspring for the next generation. Crossover involves combining genetic information from two parent individuals to create new child individuals. Mutation introduces random changes to the genetic makeup of individuals to maintain diversity in the population.

E. Population Replacement

The current population is replaced with the offspring generated through crossover and mutation. This ensures that the population evolves over generations toward better solutions.

F. Termination

The evolution process continues for a predefined number of generations. After reaching the specified number of generations, the algorithm terminates, and the best individual (solution) found throughout the evolutionary process is selected as the final solution.

GA Results and Analysis

The genetic algorithm successfully optimized the paths for our given problem, resulting in enhanced path optimization and adaptability.
The final solution obtained from the genetic algorithm demonstrated improved performance compared to alternative optimization methods.
The algorithm's parameters, including population size, mutation rate, and tournament size, were fine-tuned to achieve the best results for our specific problem.

IV. Conclusion

our study presents a comprehensive approach to optimizing sensor deployment for air pollution monitoring in urban environments. Through a combination of exploratory data analysis, optimization using Pulp, and interactive mapping, we have demonstrated the effectiveness of our methodology in enhancing air quality monitoring efficiency.
By fine-tuning parameters such as budget, coverage_threshold, and connectivity_threshold, we achieved optimal sensor deployment strategies that maximize coverage while minimizing costs and ensuring network connectivity. Our methodology's success is evident from the X-Y graphs, which illustrate the relationship between sampling fraction and trajectory similarity, and the map plots, which visualize sensor paths and pollution data distribution.
Sample results obtained during specific time intervals further validate the system's efficacy in capturing balanced datasets across multiple sensors. This balanced dataset ensures robust decision-making and accurate assessment of air quality levels throughout the monitoring period.
our approach offers a powerful tool for urban air quality management, enabling stakeholders to make informed decisions and implement targeted interventions to mitigate pollution levels effectively. As air quality continues to be a critical concern in urban areas, our methodology provides a valuable framework for optimizing sensor deployment and improving air pollution monitoring efficiency.

References

[1] Boubrima, A., W. Bechkit, and H. Rivano. *A new WSN deployment approach for air pollution monitoring*. in *2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. 2017. IEEE.

[2] Boubrima, A., W. Bechkit, and H. Rivano, On the deployment of wireless sensor networks for air quality mapping: Optimization models and algorithms. IEEE/ACM Transactions on networking, 2019. **27**(4): p. 1629-1642.

[3] Hu, S.-C., et al., *Measuring air quality in city areas by vehicular wireless sensor networks.* Journal of Systems and Software, 2011. **84**(11): p. 2005-2012.

[4] Chen, Y., et al., *A new mobile monitoring approach to characterize community-scale air pollution patterns and identify local high pollution zones.* Atmospheric Environment, 2022. **272**: p. 118936.

[5] Van den Bossche, J., et al., *Mobile monitoring for mapping spatial variation in urban air quality: Development and validation of a methodology based on an extensive dataset.* Atmospheric Environment, 2015. **105**: p. 148-161.

[6] Rashid, B. and M.H. Rehmani, *Applications of wireless sensor networks for urban areas: A survey.* Journal of network and computer applications, 2016. **60**: p. 192-219.

[7] Fascista, A., *Toward integrated large-scale environmental monitoring using WSN/UAV/Crowdsensing: A review of applications, signal processing, and future perspectives.* Sensors, 2022. **22**(5): p. 1824.

[8] Kocakulak, M. and I. Butun. *An overview of Wireless Sensor Networks towards internet of things*. in *2017 IEEE 7th annual computing and communication workshop and conference (CCWC)*. 2017. Ieee.

[9] Pavani, M. and P.T. Rao, *Urban air pollution monitoring using wireless sensor networks: A comprehensive review.* International Journal of Communication Networks and Information Security, 2017. **9**(3): p. 439-449.

[10] Ding, Y. and A.S. Fotheringham, *The integration of spatial analysis and GIS.* Computers, environment and urban systems, 1992. **16**(1): p. 3-19.

[11] Leung, Y., et al., *Integration of air pollution data collected by mobile sensors and ground-based stations to derive a spatiotemporal air pollution profile of a city.* International Journal of Geographical Information Science, 2019. **33**(11): p. 2218-2240.

[12] Trotta, A., et al., *Joint coverage, connectivity, and charging strategies for distributed UAV networks.* IEEE Transactions on Robotics, 2018. **34**(4): p. 883-900.

[13] Boubrima, A., W. Bechkit, and H. Rivano, *On the optimization of WSN deployment for sensing physical phenomena: Applications to urban air pollution monitoring.* Mission-Oriented Sensor Networks and Systems: Art and Science: Volume 1: Foundations, 2019: p. 99-145.