RESEARCH PROPOSAL

A data-driven framework for urban air quality modeling using community resiliency characteristics and machine learning algorithms

Mehdi Ashayeri, Ph.D., Assistant Professor, School of Architecture, College of Health and Human Sciences, Southern Illinois University Carbondale, mehdi.ashayeri@siu.edu

Abstract

Over half of the world's population lives in urban areas. Urban areas are major contributors to human exposure to outdoor air pollution, contributing to the global burden of diseases and death. Cities need to better understand urban air quality dynamics, and ultimately provide healthy environments for dwellers. To date, majority of urban air quality studies have incorporated environmental, co-pollutants, meteorological, and mobility factors in the model for modeling air quality of neighborhoods; however, no work has been done to assess the impacts of community resiliency characteristics for predicting intraurban air quality. Main limitations for doing so include the existence of information on human-centered factors and lack of spatial gaps on air pollution data. The growing internet of things (IoT) infrastructures and urban big data initiatives, along with advances in computational technologies, however, have provided unprecedented opportunities to predict the healthiness and well-being of neighborhoods with a higher level of accuracy.

Thus, this research aimed at developing a theoretical framework for intraurban air quality modeling ($PM_{2.5}$ concentrations as the end-point pollutant) that incorporates the human-centered resiliency factors into the urban air quality assessment workflow and localizes $PM_{2.5}$ concentrations at multiple urban scales through the application of urban big data and the Inverse Distance weighting (IDW) approach. The model will be based on the machine learning-based data-driven techniques to improve model accuracy and performance. The computation tasks will be implemented using open-sourced R programming language, and the GIS-based results will be visualized at multiple urban scales of census block and community levels by using RPubs interactive mapping platform. The Gradient Boosting Machine is selected as a robust machine learning algorithm in this research, and the coefficient of determination (R^2) will be used for performance evaluation of the model. Chicago, IL, was selected as the case study in this work for testing the proposed framework.

The proposed framework and interactive maps have the potential to aid urban health policymakers in leveraging sustainable and resilient cities. It also helps urban planners to empower analytical solutions for obtaining more realistic human health risk profiles for local developments.

Keywords: Urban air quality modeling, urban community resiliency; machine learning; IoT, urban big data; GIS mapping