A Cost-effective Framework for Indoor and Outdoor Air Quality Monitoring and Prediction

A Thesis Submitted in Partial Fulfilment of the Requirements for the degree of
Doctor of Philosophy
in
Computer Science & Engineering

by
Praveen Kumar Sharma
(NITD/PhD/CS/2017/00979)

Under the Joint Supervision of

Dr. Sujoy Saha
Assistant Professor
Department of Computer Science & Engineering, NIT Durgapur

&

Prof. Tanmay De
Professor
Department of Computer Science & Engineering, NIT Durgapur



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY DURGAPUR
DURGAPUR-713209,-INDIA
SEPTEMBER 21, 2022

Abstract

Air quality degradation is one of the major problems world wide, which draws the attention of researchers from different disciplines. The air quality quantification always becomes the prior concern of researchers followed by real-time estimation of air quality. But such monitoring highly depends upon the economic strength of a country as it requires costly set-ups and high maintenance. There are two perspective of air quality: indoor and outdoor. We are mostly familiar with the outdoor air pollution but, indoor pollution can be 5 times more polluted than that of outdoors and we spend most of our time indoors. In this work, we analysed both indoor and outdoor air quality. We have devised an end-to-end system concerning indoor as well as outdoor air quality. We developed our own customized Environment Monitoring Device (EMD) to monitor the concentration of pollutants using different types of sensors. The reliability of the developed device have been ensured using calibration technique, before real-time deployment.

We have selected classrooms for our study as students have to stay in the classroom for more than 30% of their time and indoor pollution affects physically as well as cognitively. We developed some sensor based EMDs and perform different validation as well as calibration to ensure the reliability of the EMDs. Our goal is to monitor the air quality of the building and as the behavior of pollutants in two different classrooms are different, we have to deploy an EMD in each of the classrooms. But, deployment of EMD in each classroom is not feasible as it increases the cost linearly. So, we have deployed some of the devices at few classrooms in the building and estimate the air quality of other non-monitored classrooms through Multi-layer Perceptron (MLP) and Extreme Gradient Boost (XGB) with feature interpolation technique and we achieved an accuracy of $> 91\%$. Estimation of air quality cannot help to ensure the pollution exposure so, we have opted for air quality forecasting using Long Short Term Memory (LSTM)-wF, our method outperforms other alternates of LSTM with an accuracy of $> 94\%$. We also developed an Android based interactive application to test the proposed technique in real-time as it provide alerts whenever the pollution level rises after a certain limit.

The classroom where the EMD is to be deployed, affects the estimation and forecasting. We proposed an optimal EMD placement strategy to achieve good performance of the system. The proposed technique considers the Spatio-temporal behavior of the classrooms to achieve the optimal accuracy of the system.

We also provide an estimate of the outdoor air quality in a city-scale. Though in the literature, there are work to estimate and forecast the air quality, but these techniques have several constraints which does not help the normal society without favourable infrastructure. In our work, we are mainly focused on developing a framework which can help the society in reducing the pollution exposure considering particulate matters due to its' major presence and adverse health impacts. The $PM_{2.5}$ is converted to Air Quality

Index (AQI) for getting the class-wise distribution of the air quality. We presents an end-to-end system *CiPolSe* under which we develop and deploy EMDs in 4 locations within a 23 sq. km coverage area in a suburban city. The proposed prediction model applied to the samples collected for 10 months reveals $> 75\%$ accuracy, which is better than other baseline models. The Post-Study System Usability Questionnaire (PSSUQ) feedback of the services in terms of system usability, information quality, and user interface quality reveal an overall score of 1.78, which denotes an excellent rating of the system.