Contribution ID: **52** Type: **Poster**

A machine learning practice of predicting CO2 levels with the measurement data from university departmental offices

Introduction:
In office spaces, users spend around 1/3 of their daily time, thus maintaining good air quality is an important aspect of keeping a healthy and efficient working environment. As office occupants usually have a regular daily and weekly schedule, machine learning can be a useful method to find the air quality pattern. Its result can benefit the indoor activities to improve the indoor activities and even further development on the smart buildings when the building infrastructure is ready.
In a department area of Politecnico di Milano, multiple low-cost air quality sensors have been installed since September 2023 which measure parameters mainly including temperature, humidity, radon, CO2, VOCs, air pressure and light. In this research, the CO2 data measured in the selected offices are used in this analysis practice.
This research aims to explore how machine learning can benefit long-term monitoring in improving air quality in the offices, taking CO2 as a practice, and how the existing monitoring system in the building can be improved to response to the changing indoor air quality (IAQ) level according to the model training process and the prediction performance.

Methodology:
This research tried to apply the regression learner from MATLAB to train the models based on the data collected from the selected offices respectively in the past 1 year (from 2023 Sep 05 to 2024 Sep 05). The training used the basic information as a predictor including the date of the year, time of day, room code, and day of week. Meanwhile, the level of CO2 is selected as the response in training. After the comparison of the results in RMSE (Root-mean-square Error) with different models, the training eventually selected the bagged tree model in terms of its performance and the total training time.
The trained model then was optimized by the Experiment Manager tools from MATLAB with 50 trials by tuning the 4 hyperparameters of the bagged tree model, including method, number of learning cycles, learn rate and min leaf size. The one with the best performance in RMSE was selected in the validation session.
The validation was the comparison between the measured data from Sep 06 to Oct 30 2024 and the prediction from the trained and optimized model. The performance of the validation and the model were interpreted in terms of their RMSE, residuals, and coefficient of determination (R2).

Result:
The model of rooms has RMSE results between 14.95 to 16.09 after the training and optimization varying from different rooms. Then, in comparisons between the original and model predicted CO2 values, the predicted CO2 levels basically follow the schedule of the rooms, with similar variation rates at the beginning and end of the working hours. This means that the model is able to catch the features of CO2 level variations in the selected offices based on the historical data.
However, the predictions show differences from the measurements with the dates as predictors in 2024, with RMSE from 100.52 to 107.74. These differences lay in the daily variations, especially the peaks of several days, which are much higher than the prediction results. These are due to several realistic reasons such as the number of occupants in 2024 being more than in 2023 and the schedule of occupants changing each week, etc. which are not included in monitoring and training.

Conclusion:
In general, the performance of this model currently is limited by the predictor parameters from the historical data monitored in the past 1 year, but it can already be useful in reflecting the CO2 variations in these offices.

The training process also shows the 3 types of information that could be added to the monitoring system to help benefit and respond to the IAQ level changes more smoothly and accurately, including the daily number of people, the occupancy schedule, and the ventilations by window operations.
During this training, it can be found that, in the model training for CO2 level in these offices, the number of occupants and their ventilation behaviours are the 2 influential factors that are important but not measured in the existing monitoring system, especially the number of occupants which dynamic during the year and highly related to the CO2 increasing rate and the peak level. The number of occupants and the schedule can be added as one monitoring parameter in the future to make the prediction more accurate.
On the other hand, other factors such as the dimensions of the room are less influential and can be simulated based on the calculation with the CO2 historical records and the number of occupants.
In addition, this method can be used in spaces with more occupants, such as classrooms, open offices or shopping centres with large numbers of occupants by minimizing the influence of the changes on the average number of occupants on the CO2 prediction.

Primary author: Mr YU, Yong (Design & Health Lab, Dept. Architecture, Built environment and Construction engineering, Politecnico di Milano, Milan, Italy)

Co-authors: Dr GOLA, Marco (Design & Health Lab, Dept. Architecture, Built environment and Construction engineering, Politecnico di Milano, Milan, Italy); Dr SETTIMO, Gaetano (Department of Environment and Health, Italian National Institute of Health (Istituto Superiore di Sanità), Rome, Italy); Prof. CAPOLONGO, Stefano (Design & Health Lab, Dept. Architecture, Built environment and Construction engineering, Politecnico di Milano, Milan, Italy)

Presenter: Mr YU, Yong (Design & Health Lab, Dept. Architecture, Built environment and Construction engineering, Politecnico di Milano, Milan, Italy)