Configuration and Comparative Study of Prediction Models for Indoor Air Quality

Geonhwi Lee[1], Hea-Ryeon Seo[2], and Hae-Jin Choi[3]

[1,2,3]*Department of Mechanical Engineering, Chung Ang University, Seoul, 06974, South Korea*

leegh663@cau.ac.kr
hrfighting@cau.ac.kr
hjchoi@cau.ac.kr

ABSTRACT

Since COVID-19, cultural life and working conditions have changed to be done indoors. Various harmful substances are produced indoors, and when they enter the human body through the air, they can cause serious diseases. Indoor air pollution is not visible to the naked eye, and it is not easy for people to perceive it. Human damage due to harmful indoor gases is increasing. In this study, we predict indoor air pollution occurring in daily life in advance. We collected indoor air quality data every 10 seconds from the different types of residential spaces in Seoul. For accurate prediction, we compared the prediction performances of various models, such as the ARIMA model, and the recurrent neural network (RNN) based models. In addition, the prediction performances were compared according to the size of the historical window. The comparison results revealed that for short-term interval predictions, shorter historical window sizes and simpler models were more effective. This study provides a baseline for selecting a predictive model and configuring training datasets.

1. INTRODUCTION

In recent years, COVID-19 has required people to stay indoors, which has led to people spending more time indoors than ever before. In addition to the impact of COVID-19, the development of the IT industry has enabled online shopping, telecommuting, and online meetings, which means that people are spending more time in their residential spaces than ever before. As a result, indoor air quality in residential spaces has become a growing concern.

Many pollutants are generated indoors from building materials, pets, and household activities such as cooking and cleaning. Furthermore, outdoor pollutants such as particulate matter can be brought indoors and worsen indoor air quality. These indoor air pollutants can cause serious illnesses such as lung cancer, asthma, headaches, and brain damage. According to the World Health Organization's 2019 survey of the causes of premature death around the world, indoor air pollution accounted for a high 4% of global deaths (Ritchie and Roser, 2013).

Despite the serious health risks posed by these air pollutants, many people tend to overlook them and instead focus on outdoor air quality, mistakenly believing that indoor spaces offer protection. Therefore, to reduce the human damage caused by indoor air pollution, this study predicted future indoor air quality by using historical indoor air quality data. The prediction horizon was set to 2 minutes, based on the judgment that it would be sufficient for ventilation or escaping the area. In this study, a statistical model and an artificial intelligence (AI) model were used as models for indoor air quality prediction. The performance of these models was compared by changing the conditions of the training data.

2. METHOD

2.1. Statistical model – ARIMA

In this study, the ARIMA model was used as a statistical model for indoor air quality prediction, adopting the study of "Zhang & Li (2022)". The ARIMA model is a time series prediction model that combines autoregressive (AR) and moving average (MR) models with differencing, using past values and past forecast errors to explain current values. The ARIMA model can be expressed in a formula as "Eq. (1)".

$$Y_t' = c + \sum_{i=1}^{p} \alpha_i Y_{t-i}' + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i} \tag{1}$$

In "Eq. (1)", c is the average of the difference between consecutive observations, and Y_t' is the differentiated value at time t. The ARIMA model can be applied to stationary time series data, but the collected indoor air quality data was non-stationary, so we conducted the first order of differencing. To use the ARIMA model, it is necessary to determine the values of p and q. In this study, the combination with the lowest Akaike information criterion (AIC) was selected as the optimal model.

2.2. RNN based model

Recurrent neural network (RNN) is a deep learning model that includes one or more recurrent layers in the hidden layer, where the output values from the memory cell of the hidden layer are fed as input to the next hidden layer's memory cell. Due to this characteristic, RNN is mainly used for processing sequence data. In this study, we used long short-term memory (LSTM) and gated recurrent unit (GRU). Additionally, we constructed stacked LSTM and stacked GRU models to compare the prediction performance according to model complexity. We also constructed CNN-LSTM and CNN-GRU models that preprocess time series data through a 1D convolution and a pooling before feeding the data into the LSTM and GRU models, inspired by the work of "Zhang & Li (2022)" and "Elmaz, Eyckerman, Casteels, Latre, and Hellincks (2021)".

3. Experiment

The data used in this study was collected from different residential spaces in Seoul. Environmental sensors were used to collect indoor $PM_{2.5}$, PM_{10}, CO_2, temperature, and humidity, with a sampling period of approximately 10 seconds. The data were collected for about 60 days, and this study used the 13 days of data during which the data were collected consistently.

During model training, three days of data were used as training data, one day of data was used as validation data from a total of 13 days of data, and the remaining nine days of data were used as test data. Unlike the ARIMA model, the input sequence length must be specified for the RNN-based model. In this study, we changed the length of the input sequence data from 3 to 20 minutes and evaluated the prediction performances.

Among the collected indoor air quality data, the carbon dioxide data was used in this study. Most of the models showed high prediction accuracy within a margin of error of 3%, according to the model validation results. Additionally, there was no significant difference in prediction results based on the size of the historical window, and among the RNN-based models, the models with simple structures showed better prediction performance.

4. Conclusion

In this study, we were able to accurately predict carbon dioxide concentrations through statistical models and RNN-based models. In addition, this paper provides a baseline for model selection and training data configuration in the introduction of predictive algorithms for indoor air quality monitoring. This study suggests that a simpler model with fewer historical window sizes is more effective for predicting indoor air quality in residential spaces.

Acknowledgement

This work is supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2022R1A2C2003). This research was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2022M3H4A1A02076956).

References

Hannah, R., & Max, R. (2013). Indoor Air Pollution. *Our World in Data*. Published online at OurWorldInData.org. Retrieve from: 'https://ourworldindata.org/indoor-air-pollution'

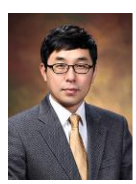
Jiaxuan, Z., & Shunyong, L. (2022). Air quality index forecast in Beijing based on CNN-LSTM multi model. *Chemosphere*. Vol. 308, part 1, 136180, doi: 10.1016/j.chemosphere.2022.136180

Furkan, E., Reinout, E., Wim, C., Steven, L., & Peter, H. (2021). CNN-LSTM architecture for predictive indoor air temperature modeling. *Building and Environment*, vol. 206, 108327. doi: 10.1016/j.buildenv.2021.108327

Biographies

Geonhwi Lee is a graduate student in the School of Mechanical Engineering, Chung-Ang University (CAU) in Seoul, Korea. He received his B.S degree in the School of Mechanical Engineering in 2023 from Chung-Ang University. His academic interests include prognostics and health management, machine learning, and signal processing.

Hea-Ryeon Seo is a graduate student in the School of Mechanical Engineering, Chung-Ang University (CAU) in Seoul, Korea. She received his B.S degree in the School of Mechanical Engineering in 2022 from Chung-Ang University. Her academic interests include mechanical system monitoring, fault diagnosis, prognostics and health management, and machine learning.



Hae-Jin Choi received M.S. and Ph.D. degrees in Mechanical Engineering from the Georgia Institute of Technology (Georgia Tech), Atlanta, GA, USA, in 2001 and 2005, respectively. He was an Assistant Professor at Nanyang Technological University, Singapore. He was also a Postdoctoral Fellow at the GWW School of Mechanical Engineering, Georgia Tech. He is currently a Professor at the School of Mechanical Engineering, Chung-Ang University, Seoul, Korea. His research interests include fault diagnosis, prognostics and health management, machine learning, simulation-based design optimization, management of uncertainty, and integrated materials and products design.