ISSN 1975-8359(Print) / ISSN 2287-4364(Online)
The Transactions of the Korean Institute of Electrical Engineers Vol. 65, No. 12, pp. 2075～2083, 2016
http://dx.doi.org/10.5370/KIEE.2016.65.12.2075

지하철 역사 실내 공기질 관리를 위한 실용적 PM_{10} 실시간 예측

A Practical Approach to the Real Time Prediction of PM_{10} for the Management of Indoor Air Quality in Subway Stations

정 갑 주*† · 이 근 영*
(Karpjoo Jeong · Keun-Young Lee)

Abstract - The real time IAQ (Indoor Air Quality) management is very important for large buildings and underground facilities such as subways because poor IAQ is immediately harmful to human health. Such IAQ management requires monitoring, prediction and control in an integrated and real time manner. In this paper, we present three PM10 hourly prediction models for such realtime IAQ management as both Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) models. Both MLR and ANN models show good performances between 0.76 and 0.88 with respect to R (correlation coefficient) between the measured and predicted values, but the MLR models outperform the corresponding ANN models with respect to RMSE (root mean square error).

Key Words : Real Time Environmental Prediction, Water Temperature, Artificial Neural Networks, Cyberinfrastructure

1. Introduction

In big cities, subways are one of the most important urban resources and infrastructure that millions of citizens use every day. In subways, the management of indoor air quality (IAQ) raises both legislative concerns and scientific interests because poor IAQ can immediately and seriously cause harmful effects on human health in underground facilities such as subways [1-6]. IAQ is defined with the measurements of various air pollutants. In Korea, there is a national law on the management of ten air pollutants in indoor air: PM_{10} (Particulate Matter with a diameter of 10 micrometers or less), CO_2 (Carbon Dioxide), Formaldehyde, TAB (Total Airborne Bacteria), CO (Carbon Monoxide), NO_2 (Nitrogen Dioxide), Radon, VOCs (Volatile Organic Compounds), Asbestos, and O_3 (Ozone). PM_{10} is usually considered to be the most common and imminent pollutant. According to this law, the concentration of PM_{10} must be kept below 150 μgm^{-3}.

The IAQ management largely depends on the ventilation systems with air filters. Ventilation systems generally dilute and displace indoor air pollutants by the mechanically-forced air circulation (often, through filters) between outdoor and indoor spaces. However, ventilation for large facilities such as subways requires a substantial amount of energy consumption and therefore is usually used when IAQ is poor. Such energy-efficient IAQ management requires IAQ to be monitored and predicted in a real time manner as shown in Fig. 1. The model-based IAQ prediction is crucial because it takes time for ventilation (i.e., air circulation) to improve IAQ and therefore the operation of ventilation systems needs to be decided, based on future IAQ values from prediction models, not on the current IAQ values from sensors.

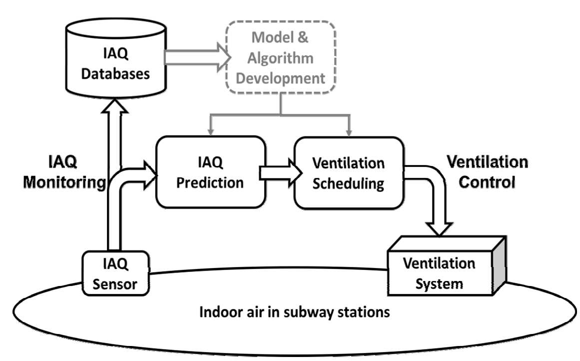


그림 1 지능형 실시간 실내 공기질 관리
Fig. 1 Intelligent real time IAQ management

† Corresponding Author : Dept. of Internet and Multimedia Engineering, Konkuk University, Korea.
E-mail: jeongk@konkuk.ac.kr
* Institute for Ubiquitous Information Technology and Applications, Konkuk University, Korea.
Received : October 31, 2016; Accepted : November 7, 2016

In this paper, we first propose a practical approach to the development of prediction models for IAQ in subways that are intended for the real time IAQ management. As opposed to conventional approaches where the research focus is generally on the model optimization for high precision prediction, our approach aims at the effective applicability of prediction models in real world environments. Second, we present multiple prediction models for PM_{10} that are designed according to the model design strategies. Then, we evaluate their performances. They are implemented as Artificial Neural Network (ANN) and Multiple Linear Regression (MLR) models. These models are developed on the datasets from a large scale project to monitor the IAQ of subway stations in Seoul from 2008 to 2012. In the project, the concentration of PM_{10} was measured by TMS (Tele-Monitoring System) every 30 seconds and so 120 measurements were collected during each hour. In fact, these datasets have been used for other studies on IAQ [24-31].

This current work is a part of our development project for ICT systems to integrate monitoring, data management, prediction, and control for smart city applications in an intelligent and real time manner [14, 15, 20-22]. The ICT systems include a sensor network system for real time monitoring, a sensor data management system, a simulation support system for various prediction models, and a planning system for control systems.

2. Challenges to the IAQ prediction in Subways and Design Strategies for Prediction Models

In this paper, we explain challenges for the effective prediction of IAQ in subways and then propose a practical approach to how to design prediction models to address those challenges efficiently. The challenges are as follows. First, IAQ sensors are prone to the easy contamination of air pollutants such as dusts, but in subways, indoor air contains a lot higher concentrations of pollutants than ambient air. Without frequent and careful maintenance (e.g., cleaning and calibration), it is difficult to expect sensors to produce high precision measurement values constantly. Once they are contaminated by high concentrations of such pollutants, sensors continue to produce incorrect measurement values until they are cleaned and calibrated again. However, such sensor contamination is, unfortunately, often unavoidable because sensors are always exposed to air pollutants.

Second, the effective real time IAQ monitoring crucial for the real time prediction depends on a number of very expensive IAQ sensors and high quality maintenance work on those sensors. In subways, indoor air occupies a number of large spaces such as platforms and tunnels and a single point IAQ measurement cannot represent the IAQs of all the spaces. For this reason, the construction and operation of an effective real time IAQ monitoring system requires both a large amount of investment and a substantial amount of maintenance work.

Finally, when raw datasets directly from sensors contain lots of incorrect data, the development of prediction models generally depends on both a large amount of data preprocessing and sophisticated optimization in order to increate prediction performances. Although they may improve prediction performances against both training and test datasets, such preprocessing and optimization efforts may also increase the risk of overfitting significantly where such efforts make prediction models too customized for the training datasets but unsuitable for a variety of new data from real world environments.

In order to address such challenges, we propose a practical approach to the development of prediction models for subways as follows:

- *Single pollutant-based prediction.* PM_{10} is used as the only indicator for the overall IAQ management. In other words, only one type of an IAQ sensor is used for the IAQ monitoring. This single pollutant-based approach is intentionally taken for three reasons. First, PM_{10} is considered to be the most imminent threat. Second, the concentration levels of such air pollutants as PM_{10}, CO, CO_2, SO_2, and NO_2 tend to be positively co-related in most cases [4]. That is, if the concentration of PM_{10} is higher, then it is very likely that the concentrations of the other pollutants are also higher. Finally, IAQ sensors are usually costly and require frequent and intensive maintenance. Therefore, the monitoring of a fewer pollutants at run time usually cause much less development costs and maintenance overheads.
- *One-hour data based prediction.* For the fast system initialization or failure recovery, the data (in fact, 120 measurements) collected only during the last one hour is used for each prediction. Although the use of more data during a longer period of time may improve prediction performances, such prediction models require longer time for the system initialization and recovery from various failures.
- *Hourly prediction on three representative values.* Three statistical representative values computed from 120 measurements during each hour (i.e., with sensor sampling at each 30 seconds) are used to represent the hour's PM_{10} concentration: the hourly maximum, average, and minimum values. In other research work on IAQ, only average values are usually used [24-31]. The motivation for three

representative values is twofold. First, those three values as the input variables are more representative. Second, these representative values as the output variables (i.e., prediction results) allow more variety of ventilation control and other IAQ management. For example, the hourly maximum prediction can be used when more priority is on IAQ in ventilation control, but the hourly minimum prediction when more priority is on efficient energy consumption.

3. Theory

3.1 Multiple Linear Regression

Multiple linear regression (MLR) is one of the statistical techniques to minimize the residual sum of squared errors between the measured and predicted values [32]. MLR has been used for predicting PM10 concentrations in ambient air [33]. Recently, MLR has also been used for PM10 in subways [24, 30, 31]. The equation for the general MLR is:

$$y=\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n$$

where y, $x_i \in R$, and β_i denote the output variable, the i-th variable of the input vector $X \in R_{n+1}$, and the i-th regression coefficient in the model, respectively. The linear regression model of a system with one output is built by optimizing the function:

$$\min_\beta \sum_{k=1}^{N} |\beta^T X_k - y_k|^2$$

where $X_k \in R_{n+1}$, $y_k \in R$, and $\beta \in R_{n+1}$ are the k-th input vector for the model, the k-th measured value, and the vector of the weighting coefficients for n= the number of features in X_k and N=the number of data samples.

3.2 Artificial Neural Network

An artificial neural network (ANN) basically consists of three types of node (or, neuron) layers: a single input layer, multiple hidden layers and a single output layer. Each node of each layer can be connected to those of preceding and succeeding layers, but not to those of the same layer. Such links between nodes have modifiable connection weights. In addition, each node of every layer except the input layer is connected to a single bias node of its preceding layer. The numbers of the layers and the nodes of each layer need to be optimized [34-36].

In this paper, the ANN with a single hidden layer is used. Fig. 2 presents the structure of a three-layer perceptron network consisting of an input layer with N_I nodes, a single hidden layer with N_H nodes, and an output layer with one node. The w_{ij} and v_j are the connection weights from the *i-th* node of the input layer to the *j-th* node of the hidden layer and from the *j-th* node of the hidden layer to the output layer node, respectively. The initial values for w_{ij} and v_j are randomly generated in [-1,1] at the beginning. The total input signal for the *j-th* node of the hidden layer is:

$$u_j=\sum_{i=1}^{N_I} x_i w_{ij}+b_j$$

where x_i is the *i-th* element value of the input vector and b_j is the bias for the *j-th* node of the hidden layer. Each node of the hidden layer emits an output signal through the activation function $f(u_j)$ that is,

$$y_j=f(u_j).$$

A number of differentiable activation functions can be applied to the ANN, but in this paper, the following sigmoid function is used:

$$f(u)=1/(1+e^{-u}).$$

The input signal for the output node is:

$$s=\sum_{j=1}^{N_H} y_j v_j+b$$

where b is the bias. The node of the output layer uses a linear activation function to emit the same output signal as an input signal:

$$g(s)=s.$$

The ANN is trained by a backpropagation algorithm with gradient descent and momentum terms that minimize the network error function [37, 38]. Refer to [39] for a further study of the ANN and the backpropagation algorithm.

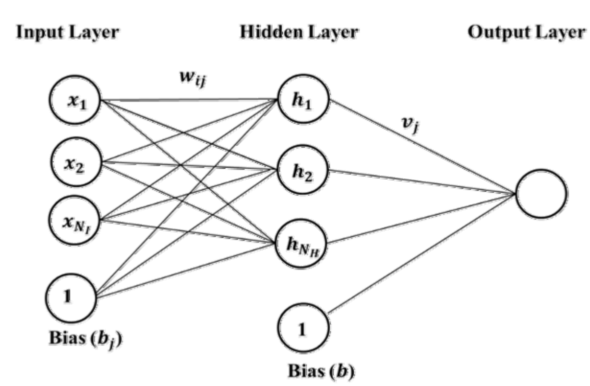


그림 2 단일 숨김 계층을 가진 인공신경망 구조
Fig. 2 ANN structure with a single hidden layer

4. Materials and Methods

4.1 Data Sampling, Cleansing, and Preprocessing

The datasets used for the development of prediction models in this paper were collected at subway stations in Seoul from 2008 to 2012. These stations are anonymous because of their administration policy. The project to collect those datasets was a joint research effort by the research teams from Konkuk University, Kyeong Hee University, University of Seoul, and some industrial companies. The concentration of PM_{10} was measured by TMS (Tele-Monitoring System) every 30 seconds and so 120 measurements were collected during each hour. In fact, the research team at Kyeong Hee University has been carrying out a number of IAQ studies on these datasets [24-31]. For this paper, the hourly maximum, average, and minimum values were computed from the raw data collected between Jan. 2011 and Oct. 2012.

For various reasons such as contamination, IAQ sensors sometimes fail to operate correctly. Data from such faulty sensors is very likely to be incorrect and such incorrect data should be excluded from analyses. Two simple data validation techniques were used to discard incorrect data. First, those measured values of PM_{10} that were greater than 1,000 μgm^{-3} were assumed to be from faulty sensors and therefore removed. Second, sensors that continued to generate zero values during more than two hours were assumed to be faulty. The data values from those sensors during such periods of time were deleted.

As a result, the dataset of this study consisted of 9,780 records. The dataset was divided into the training and test datasets. The 8,004 observations between 1/1/2011 and 4/22/2012 are used to train prediction models: the training dataset. The remaining 1,776 observations between 4/23/2012 and 10/15/2012 were used to evaluate the performances of prediction models: the test dataset. Table 1 shows some basic statistical analyses of the training and test datasets.

4.2 The proposed method

4.2.1 The design of prediction models

In this paper, we develop three PM_{10} prediction models designed as follows:

- Every model uses three variables as the input vector: the hourly maximum, average, and minimum of PM_{10} concentrations (in fact, 120 data samples) from time t-1 to time t. These input variables are represented to be Max_{t-1}, Avg_{t-1} and Min_{t-1}, respectively. That is, only the PM_{10} concentrations during the last one hour are used for prediction.
- We design three 1-hour-ahead prediction models that predict the hourly maximum, average, and minimum of PM_{10} concentrations from time t to time t+1, respectively. That is, each prediction model predicts one of three statistical representatives of the PM_{10} concentrations during the next one hour. The output variables from three prediction models are represented as Max_t, Avg_t and Min_t, respectively.
- However, all of these output variables are simply denoted as the common variable $PM_{10}(t)$ in model equations unless it is required to distinguish these output variables explicitly. Therefore, a single model equation including the variable $PM_{10}(t)$, in fact, represents a number of prediction models.
- Finally, the MLR and ANN modeling techniques are separately applied to implement the three prediction models so far explained. Therefore, the total six prediction models (i.e., three MLR models and three ANN models) are presented in this paper. Hereafter, three MLR models are represented simply as MLR-Max_t, MLR-Min_t, MLR-Avg_t and three ANN models as ANN-Max_t, ANN-Min_t, ANN-Avg_t.

표 1 학습 데이터셋과 테스트 데이터 셋의 평균, 표준편차, 범위

Table 1 Means, Standard deviatios and ranges of the training and test datasets

	Mean	Standard deviation (SD)	Range	Numbers of samples
Training data	66.02	73.97	0.9-615.02	8004
Test data	50.16	35.38	0-251.11	1776

4.2.2 Multiple Linear Regression Models

The regression coefficients for the following MLR equation are computed:

$$PM_{10}(t) = \beta_0 + \beta_1 Avg_{(t-1)} + \beta_2 Max_{(t-1)} + \beta_3 Min_{(t-1)}.$$

As explained in Section 4.2.1, this equation, in fact, represents three models. Table 2 shows the regression coefficients of each prediction model computed from the training dataset.

표 2 3가지 다중회귀분석 모델에 대한 회귀계수

Table 2 Regression coefficients for three MLR models

Model	β_0	β_1	β_2	β_3
MLR-Avg_t	5.13	1.09	-0.070	-0.09
MLR-Max_t	15.27	1.61	0.070	-0.66
MLR-Min_t	3.11	0.02	-0.006	0.89

4.2.3 Artificial Neural Network Models

In this study, the ANN models are designed as a feed-forward network with a single hidden layer. As explained in Section 3.2, they use the sigmoid activation function in the hidden layer and the linear activation function in the output node. A single hidden layer is used according to Bishop's study [40] to indicate that multiple hidden layers do not improve model performances significantly in most cases. Fig. 3 shows the structure of the ANN models that has one input layer with three input variables and one additional node accounting for bias, one hidden layer with one additional bias node, and one output variable of the output layer.

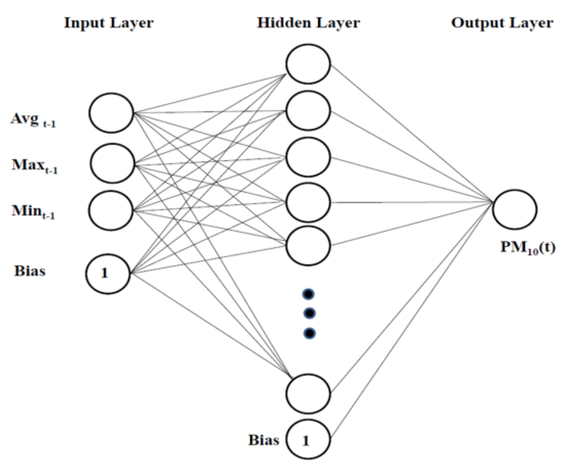


그림 3 PM_{10} 예측에 대한 인공신경망 모델의 구조
Fig. 3 Structure of the ANN Models for the PM_{10} prediction

표 3 인공신경망 모델들의 파라미터
Table 3 Parameters of ANN models

Model	Hidden Nodes	Learning Rate	Best MSE
ANN-Avg_t	25	0.25	0.038
ANN-Max_t	25	0.25	0.610
ANN-Min_t	22	0.35	0.063

The ANN models are built as follows. All the model weights are randomly initialized in [-1,1] at the beginning. The set of parameters for building the ANN models include the learning rate, the number of nodes in the single hidden layer, and the maximum number of training epochs [41]. The performances of the ANN models are evaluated and compared for the numbers of nodes in the hidden layer from 5 to 25 and for the learning rates from 0.01 to 1.0 in the increment of 0.05. The performance is evaluated by the mean square error (MSE) between the model output and the measured data. Table 3 shows the values for the parameters when each ANN model shows the best performance.

Generalization is crucial for prediction models in order to avoid the over-fitting problem that a model is too dependent on its training data [42]. For this purpose, the early stopping technique is employed. The training dataset is randomly split into two sets. 80% of the training dataset is used for model training. The remaining 20% of the training dataset (called the validation dataset) is applied to model testing. The training process is stopped when the network starts to over-fit the data, that is, the error value (i.e., MSE) begins to increase for the validation set, again.

4.3 Performance Evaluation

The performance evaluation of prediction models employs two techniques: the root mean square error (RMSE) and correlation coefficient (R). RMSE shows residual errors between the measured and predicted values. R indicates the strength and the direction of a linear relationship between measured and predicted values. These are defined as:

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i-O_i)^2}$$

$$R=\frac{\sum_{i=1}^{n}(O_i-\overline{O})(P_i-\overline{P})}{\sqrt{\sum_{i=1}^{n}(O_i-\overline{O})^2\sum_{i=1}^{n}(P_i-\overline{P})^2}}$$

where $O_i, P_i, \overline{O}, \overline{P}$, and n denote are a measured value, a predicted value, the average of all the measured values, the average of all the predicted values, and the number of all the data samples, respectively.

5. Result and Discussion

With respect to R and RMSE, Table 4 shows the performance comparison of all the six 1-hour-ahead prediction models (MLR-Max_t, MLR-Min_t, MLR-Avg_t, ANN-Max_t, ANN-Min_t, ANN-Avg_t). First, both MLR and ANN models show good performances between 0.76 and 0.88 with respect to R between the measured and predicted values, although the MLR models perform a little better than the ANN models for the prediction of the hourly average and minimum values. That is, all the models have strong linear relationships between the measured and predicted data values.

표 4 Mean, SD, RMSE, R에 관련한 성능평가 결과

Table 4 Performance evaluation results with respect to Mean, SD, root mean square error (RMSE), and correlation coefficients (R).

Model	Mean	SD	RMSE	R
MLR-Avg$_t$	51.20	32.80	16.10	0.88
MLR-Max$_t$	77.58	42.16	32.33	0.76
MLR-Min$_t$	36.65	26.77	12.96	0.89
ANN-Avg$_t$	54.38	56.16	31.68	0.86
ANN-Max$_t$	82.20	110.86	80.02	0.76
ANN-Min$_t$	52.35	47.21	32.05	0.84

However, the MLR models show better performances than the ANN models in terms of RMSE between the measured and predicted values. That is, the MLR models have much less prediction errors than the ANN models. In fact, the mean of the predicted dataset from MLR models (e.g., 51.2 for MLR-Avg$_t$) is closer to that of the test dataset (i.e., 50.16) than that of the predicted dataset from ANN models (i.e., 54.38 for ANN-Avg$_t$). This means that both the MLR and ANN models have almost the same linearity between the measured and predicted values, but in the MLR models, actual predicted values are closer to measured values. For MLR-Avg$_t$ and ANN-Avg$_t$, Fig. 4 shows and compares both the measured and predicted values, respectively.

5. Related Work

There have recently been a number of studies on IAQ in subways, but a limited amount of research work on the IAQ prediction for real time ventilation control [1, 2, 3, 23-31]. Among them, a research team at Kyung Hee University has been most actively and comprehensively carrying out various analysis and model development studies on IAQ in subways for the last few years [24-31]. In [25-29], they presented statistical IAQ models for monitoring, analyses and diagnoses, but not for prediction and control. In [30], they developed two multivariate daily prediction models for PM_{10} and $PM_{2.5}$ that use multiple air pollutants as input variables. One model was based on Multiple Linear Regression (MLR) and another model on Artificial Neural Networks (ANN). Their performances were compared and the MLR model outperformed the ANN model. More recently, they developed hourly prediction models (based on MLR) for ventilation control [31]. In this work, they also studied both energy consumption issues and the effects of ambient air quality on IAQ. In addition, they also looked into how to optimize the control of the ventilation systems.

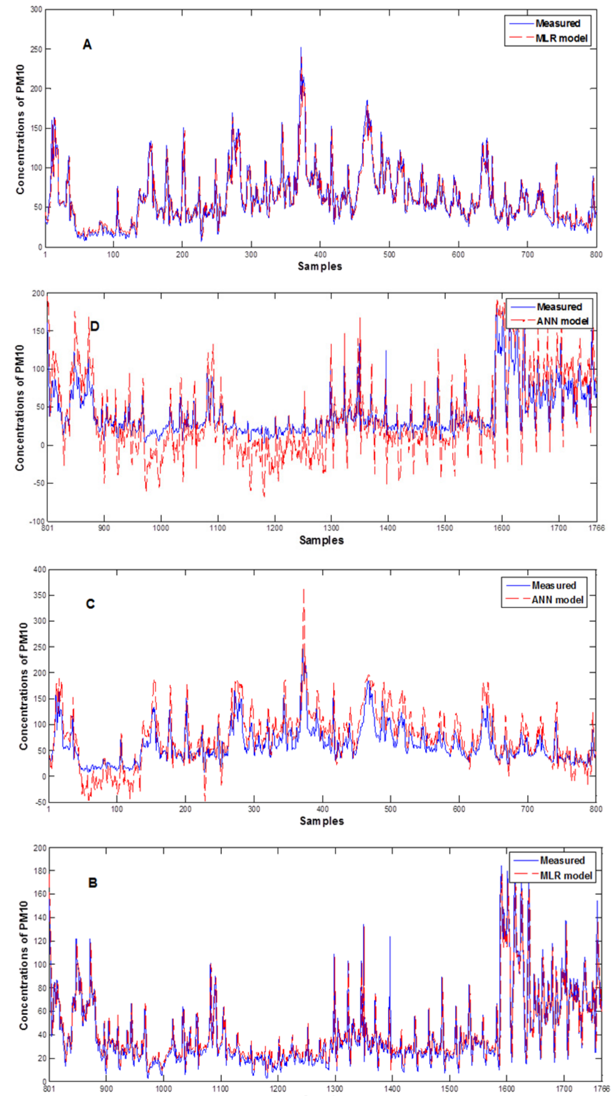


그림 4 1시간 평균 예측에 대한 측정값과 예측값 비교.

Fig. 4 Comparison between the measured and predicted values for the hourly average prediction.

There are also other research efforts to study IAQ in subways. In [4], the correlations between multiple air pollutants were studied and the concentration levels of such air pollutants as PM_{10}, CO, CO_2, SO_2, and NO_2 were shown to be positively co-related in most cases. In [1, 2, 23], the IAQ of subways in Prague, Hong Kong, and Los Angeles was studied, respectively. Their studies ware aimed only at monitoring and analyses.

As opposed to our research work in this paper explicitly aimed only at hourly prediction for intelligent, real time management of IAQ in subways, those other research efforts except [31] were aimed at only monitoring or daily prediction and therefore, in practice, inapplicable to real time ventilation control. In [31], hourly prediction models for PM_{10} and $PM_{2.5}$ were studied. However, those models required monitoring multiple air pollutants (i.e., more runtime maintenance overheads).

6. Conclusion

In large building and underground facilities such as subways, poor IAQ can cause seriously harmful effects on human health. The intelligent, real time IAQ (Indoor Air Quality) management where IAQ is monitored, predicted and controlled in a real time manner is very important. However, there has been a very limited amount of research work on the IAQ prediction for the real time management of IAQ in subways because such research requires a substantial amount of IAQ monitoring data and the monitoring of IAQ is still very challenging.

In this paper, we first raised challenges for the development of IAQ prediction models in subways and proposed a model design approach intended to address such challenges effectively. The challenges include the reliability of IAQ sensors, the construction and operation of IAQ monitoring systems, and the risk of overfitting. With those challenges, we explained how conventional model development approaches could be inefficient. The proposed design approach aims at the effective applicability of prediction models in real world environments.

Second, we presented three prediction models for PM_{10} whose designs are based on the model design approach. Finally, we evaluated their performances. The three prediction models were implemented by both Multiple Linear Regression (MLR) and Artificial Neural Networks (ANN), respectively. As a result, the total six models were developed and their performances were compared with respect to both RMSE (Root Mean Square Error) and R (correlation coefficient).

Overall, the MLR models outperformed the corresponding ANN models. That is, the MLR modeling method was shown to be more effective for the prediction of PM_{10} than the ANN method.

In the design of these multiple hourly prediction models, we intended such prediction models to be easily integrated into ICT (Information and Communications Technology) systems for real time monitoring and control [14, 15, 20-22]. Because of such motivations, the design of these prediction models are explicitly intended to reduce runtime system operation and maintenance overheads.

감사의 글

This work was funded by the Weather Information Service Engine (WISE) project under Grant KMIPA-2012-0001-1.

References

[1] Branis M. The contribution of ambient sources to particulate pollutions in spaces and trains of the Prague underground transport system, Atmospheric Environment 2006, 40, 348-356.

[2] Kam W, Cheung K, Daher N, Sioutas C. Particulate matter (PM) concentrations in underground and ground-level rail systems of the Los Angeles metro, Atmospheric Environment 2011, 45, 1506-1516.

[3] Kim MJ, Braatz RD, Kim JT, Yoo CK. Indoor air quality control for improving passenger health in subway platforms using an outdoor air quality dependent ventilation system, Building and Environment 2015, 92, 407-417.

[4] Song HB, Kim DK, Do HS, Han GH. Leem JK, Hong SH, Kim MH, Baek SO. Evaluation of indoor air quality and the perception of subway in Taegu area, Journal of the Korean Society of Environmental Engineers 2001, 23, 337-348.

[5] Cheng YH, Yi LL, Liu CC. Levels of PM10 and PM2.5 in Taipei rapid transit system, Atmospheric Environment 2008, 42, 7242-7249.

[6] Carten? A, Furi C, Stefano C. Underground and ground-level particulate matter concentrations in an Italian metro system, Atmospheric Environment 2015, 101, 328-337.

[7] Liu H, Lee S, Kim M, Shi H. Multi-objective optimization of indoor air quality control and energy consumption minimization in a subway ventilation system, Energy Build. 2013, 66, 553-561.

[8] Kitchin R. The real-time city? Big data and smart urbanism, GeoJournal 2014, 79, 1-14.

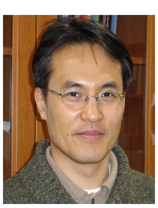
[9] Su K, Jie L, Hongbo F. Smart city and the applications, 2011 International Conference on Electronics, Communications and Control (ICECC), 1028-1031.

[10] Agarwal Y, Balaji B, Gupta R, Lyles J, Wei M, Weng T. Occupancy-driven energy management for smart building automation, Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy- Efficiency in Building 2010.

[11] Chen H, Chou P, Duri S, Lei H, Reason J. The design and implementation of a smart building control system, 2009 IEEE International Conference on e-Business Engineering, 255-262.
[12] Baheti R, Helen G. Cyber-physical systems, The impact of control technology 2011, 12, 161-166.
[13] Rajkumar R, Lee I, Sha L, Stankovic J. Cyber-physical systems: the next computing revolution, In Proceedings of the 47th Design Automation Conference ACM, 731-736.
[14] Lee J, Lee KY, Jeong K, Jiang M, Kim B, Hwang S. A Cyberinfrastructure-based Approach to Real Time Water Temperature Prediction, PRAGMA Workshop on International Clouds for Data Science 2015.
[15] Ngoc TQ, Lee J, Gil KJ, Jeong K, Lim SB. An ESB based micro-scale urban air quality monitoring system, 2010 IEEE Fifth International Conference on Networking, Architecture, and Storage, 288-293.
[16] Atzori L, Antonio I, Giacomo M. The internet of things: A survey, Computer networks 2010, 54, 2787-2805.
[17] Wei C, Li Y. Design of energy consumption monitoring and energy-saving management system of intelligent building based on the internet of things, Electronics, Communications and Control (ICECC), 2011 International Conference on.
[18] So AT, Chan WL. Intelligent building systems, Springer Science & Business Media 2012, 5.
[19] Kim J, Chu C, and Shin S. ISSAQ: An integrated sensing systems for real-time indoor air quality monitoring, IEEE Sensors Journal 2014, 14.12, 4230-4244.
[20] Joe W, Lee J, Jeong K. CSN: the conceptually manageable sensor network, International Journal of Distributed Sensor Networks 2015.
[21] Jiang M, Lee J, Jeong K. A Data Stream-Based, Integrative Approach to Reliable and Easily Manageable Real Time Environmental Monitoring, International Journal of Distributed Sensor Networks 2015.
[22] Cui Z, Jiang M, Jeong K, Kim B. A cloud database service approach to the management of sensor data, 2014 International Conference on Information Science & Applications (ICISA).
[23] Lau J, Hung WT, Cheung CS. Interpretation of air quality in relation to monitoring station's surroundings, Atmospheric Environment 2009, 43, 769-777.
[24] Kim M, SankaraRao B, Kang O, Kim J, Yoo CK. Monitoring and prediction of indoor air quality (IAQ) in subway or metro systems using season dependent models, Energy Build. 2012, 46, 48-55.
[25] Kim YS, Kim MJ, Lim JJ, Kim JT, Yoo CK. Predictive monitoring and diagnosis of periodic air pollution in a subway station, Journal of Hazardous Materials 2010, 183, 448-459.
[26] Kim YS, Kim MJ, Kim IW, Yoo CK. Multivariate monitoring and local interpretation of indoor air quality in Seoul's metro system, Environmental Engineering Science 2010, 27, 721-731.
[27] Oh TS, Kim MJ, Lim JJ, Kang OY, Shetty KV, SankaraRao B, Yoo CK, Park JH, Kim JT. A real-time monitoring and assessment method for calculation of total amounts of indoor air pollutants emitted in subway stations, Journal of the Air & Waste Management Association 2012, 62, 517-526.
[28] Lee S, Liu H, Kim MJ, Kim JT. Yoo CK. Online monitoring and interpretation of periodic diurnal and seasonal variations of indoor air pollutants in a subway station using parallel factor analysis (PARAFAC). Energy Build. 2014, 68, 87-98.
[29] Kim YS, Kim IW, Kim JC, Yoo CK. Multivariate monitoring and local interpretation of indoor air quality in Seoul's metro system. Environ. Eng. Sci. 2010, 27, 721-731.
[30] Kim MH, Kim YS, Lim JJ, Kim JT, Sung SW, Yoo CK. Data-Driven Prediction Model of Indoor Air Quality in an Underground Space, Korean J. Chem. Eng. 2010, 27, 1675-1680.
[31] Lee S, Kim M, Kim J, Yoo CK. In search for modeling predictive control of indoor air quality and ventilation energy demand in subway station, Energy Build. 2015, 98, 56-65.
[32] Burger EM, Moura SJ. Gated ensemble learning method for demand-side electricity load forecasting, Energy Build. 2015, 109, 23-34.
[33] Shahraiyni HT, Sodoudi S. Statistical Modeling Approaches for PM10 prediction in urban area; A review of 21st-century studies, Atmosphere 2016, 7
[34] Maier HR, Dandy GC. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental Modelling and Software 2000, 15, 101-124.
[35] Sahoo GB, Ray C. Flow forecasting for a Hawaiian stream using rating curves and neural networks. Journal of hydrology 2006, 317, 63-80.
[36] Sahoo GB, Ray C. Predicting flux decline in cross-flow membranes using artificial neural networks and genetic algorithms. Journal of Membrane Science 2006, 283, 147-157.
[37] Haykin S. Neural Networks: A Comprehensive Foundation. Macmillan, New York 1999.

[38] ASCE Task Committee, Artificial neural network in hydrology. Journal of Hydrologic Engineering 2000, 5, 124-144.
[39] Sahoo GB, Schladow SG, Reuter JE. Forecasting stream water temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models. Journal of Hydrology 2009, 378, 325-342.
[40] Bishop C. Neural Networks for Pattern Recognition. Oxford University Press, New York 1995.
[41] Hill DJ, Minsker BS. Anomaly detection in streaming environmental sensor data: A data-driven modeling approach. Environmental Modeling and Software 2010, 25, 1014-1022.
[42] Shilink U, Doring S, Pelikan E, Nunnari G, Cawley G, Junninen H, Greig A, Foxall R, Eben K, Chatterton T, Vondracek J, Richter M, Dostal M, Bertucco L, Kolehmainen M, Doyle M. A rigorous inter-comparison of ground-level ozone predictions. Atmospheric Environment 2003, 37, 3237-3253.
[43] Chaloulakou A, Grivas G, Spyrellis N. Neural network and multiple regression models for PM10 prediction in Athens: A Comparative assessment, Journal of the Air & Waste Management Association 2003, 53, 1183-1190.

저 자 소 개



정 갑 주 (Karpjoo Jeong)
1962년 9월생. 1984년 서울대학교 공과대학 컴퓨터공학과 학사 졸업. 1986년 동 대학교 대학원 컴퓨터공학과 석사 졸업. 1996년 New York University Computer Science 학과 박사 졸업. 1997년~현재 건국대학교 교수(인터넷미디어공학과)

이 근 영 (Keun-young Lee)
1979년 3월생. 2003년 서강대학교 경제학과 졸업. 2009년 KAIST 수리과학과 박사. 2010~2013년 건국대학교 신기술융합과 BK21 연구 교수. 2013~현재 건국대학교 유비쿼터스 정보 기술 연구원 연구 교수