Journal of Hazardous Materials 305 (2016) 209–218

Contents lists available at ScienceDirect

Journal of Hazardous Materials

journal homepage: www.elsevier.com/locate/jhazmat

A robust localized soft sensor for particulate matter modeling in Seoul metro systems

Hongbin Liu [a,b], ChangKyoo Yoo [b,*]

[a] Jiangsu Provincial Key Lab of Pulp and Paper Science and Technology, College of Light Industry Science and Engineering, Nanjing Forestry University, Nanjing 210037, China
[b] Department of Environmental Science and Engineering, College of Engineering, Kyung Hee University, Yongin 446701, South Korea

HIGHLIGHTS

- A robust localized soft sensor was proposed to predict particulate matter ($PM_{2.5}$) in a subway.
- The just-in-time (JIT) learning-based methods outperformed global methods in predicting $PM_{2.5}$.
- An outlier detection step was integrated into the local soft sensor to improve its robustness.
- The prediction performance of JIT-LSSVR was greatly improved in comparison to that of global LSSVR.

ARTICLE INFO

Article history:
Received 19 August 2015
Received in revised form
24 November 2015
Accepted 25 November 2015
Available online 2 December 2015

Keywords:
Just-in-time learning
Indoor air quality
Least squares support vector regression
Partial least squares
Particulate matter
Soft sensors
Subway systems

ABSTRACT

Developing accurate soft sensors to predict and monitor the indoor air quality (IAQ) of hazardous pollutants that accumulate in underground metro systems is of key importance. The just-in-time (JIT) learning technique possesses a local feature that can track the variations in the dynamic process more effectively, which is different from the traditional soft sensor modeling methods, such as partial least squares (PLS), which models the process in an offline and global way. In this study, a robust soft sensor that combined the JIT learning technique with a least squares support vector regression (LSSVR) method, named JIT–LSSVR, was derived in order to improve the prediction performance of a $PM_{2.5}$ soft sensor in a subway station. Additionally, in order to eliminate the adverse effects caused by the outliers in the process variables, an outlier detection step was integrated into the JIT–LSSVR modeling procedure. The performance evaluation results demonstrated that the proposed robust JIT–LSSVR soft sensor has the capability to model nonlinear and dynamic subway systems. The root mean square error of the JIT–LSSVR soft sensor was improved by 55% in comparison with that of the LSSVR soft sensor.

1. Introduction

Metro systems have been considered to be an important and popular mode of transportation because they relieve congestion and provide an efficient solution to the problem of insufficient public transportation. Given the fact that more than eight million commuters utilize the Seoul metropolitan subway system daily, the indoor air quality (IAQ) in metro systems, particularly in underground subway stations, is very important to public health and has attracted a great deal of public attention [1]. In the case of the Seoul metropolitan subway system, the Korean Ministry of Environment has established regulations in order to monitor and control the levels of several hazardous pollutants, such as carbon monoxide (CO), carbon dioxide (CO_2), nitrogen dioxide (NO_2), particulate matters with diameters less than 10 (PM_{10}) and 2.5 μm ($PM_{2.5}$), airborne fungi, and carbonyl compounds. Excessive exposure to these hazardous pollutants, among which particulate matter is thought to be more toxic than the others, can cause serious health consequences, such as cardio-respiratory illnesses and mortality [1–6]. In particular, the concentration of particulate matter, including PM_{10} and $PM_{2.5}$, in underground subway stations is higher than in outdoor spaces in Seoul [7]. Therefore, developing accurate prediction models of particulate matter for monitoring the IAQ in subway stations has become an important issue of public concern.

Several prediction and monitoring techniques have been conducted in order to ensure passengers' health in metro systems [8–11]. Kim et al. [8] compared the prediction performances of PM_{10} and $PM_{2.5}$ in a subway station using three data-driven predic-

* Corresponding author. Fax: +82 31 202 8854.
E-mail address: ckyoo@khu.ac.kr (C. Yoo).

http://dx.doi.org/10.1016/j.jhazmat.2015.11.051
0304-3894/

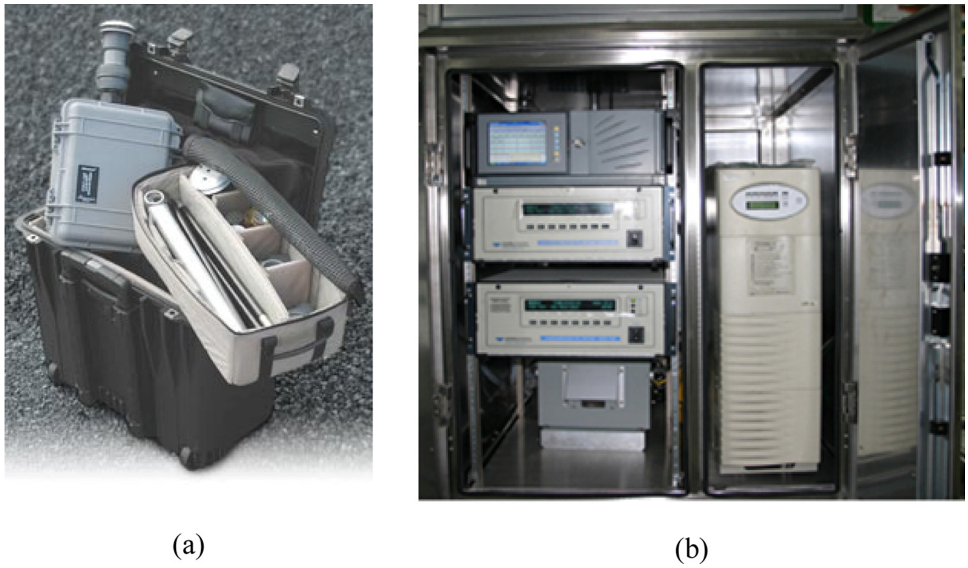

Fig. 1. Measuring equipment used for IAQ data collection: (a) mini-volume air sampler and (b) TMS system.

tion models, in which the recurrent neural networks method was more appropriate than the multiple linear regression and neural networks methods. Kim et al. [9] developed a predictive fault diagnosis system based on a multi-way principal component analysis in order to monitor the periodic patterns of the air pollutants in a subway system. Kim et al. [10] proposed season dependent models for the monitoring and prediction of PM_{10} and $PM_{2.5}$ in a subway station. The season dependent models utilized the partial least squares (PLS) modeling method for prediction purposes. In the work of Liu et al. [11], a recursive partial least squares (RPLS) method, which is widely used to model dynamic processes, was used in order to improve the prediction performance of the indoor air quality soft sensors in an underground subway station.

Soft sensors are a key technique used to estimate reliable product quality or other important variables when online analyzers are not available or encounter failures. These sensors have been widely used in many process industries [12]. Taking subway systems as an example, IAQ sensors sometimes exhibit poor quality and low reliability due to the long term usage and the hostile underground environment. Therefore, some sensors may experience failures, including bias, drifting, complete failure, and precision degradation [13]. In this case, soft sensors are good alternatives to the faulty real sensors. According to Kadlec et al. [14], data-driven soft sensors can be categorized into statistical method-based and soft computing-based sensors. In the last few decades, linear statistical modeling methods, such as PLS and principle component regression (PCR), have been extensively used for soft sensors. On the other hand, nonlinear soft sensors, which can capture the nonlinear characteristics of the process variables, have already been intensively researched; examples of such soft sensors include the artificial neural network (ANN) [15], ANN-based PLS or PCR, support vector machine (SVM) [16], and adaptive neuro-fuzzy inference

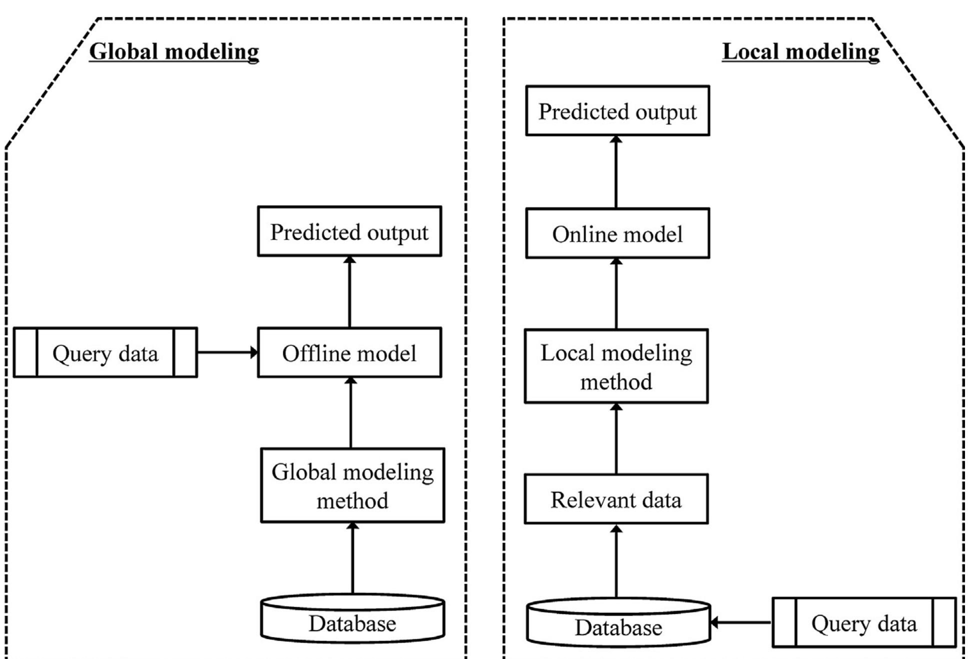

Fig. 2. Comparison between the global and local modeling methods.