Applied Energy 87 (2010) 2087–2094

Contents lists available at ScienceDirect

Applied Energy

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

Virtual models of indoor-air-quality sensors

Andrew Kusiak *, Mingyang Li, Haiyang Zheng

Department of Mechanical and Industrial Engineering, 3131 Seamans Center, The University of Iowa, Iowa City, IA 52242 – 1527, United States

ARTICLE INFO

Article history:
Received 22 September 2009
Received in revised form 23 November 2009
Accepted 8 December 2009
Available online 12 January 2010

Keywords:
Sensor modeling
Heating
Ventilation
Air conditioning systems
Sensor monitoring
Data mining
Neural networks
Indoor air quality
Statistical control charts

ABSTRACT

A data-driven approach for modeling indoor-air-quality (IAQ) sensors used in heating, ventilation, and air conditioning (HVAC) systems is presented. The IAQ sensors considered in the paper measure three basic parameters, temperature, CO_2, and relative humidity. Three models predicting values of IAQ parameters are built with various data mining algorithms. Four data mining algorithms have been tested on the HVAC data set collected at an office-type facility. The computational results produced by models built with different data mining algorithms are discussed. The neural network (NN) with multi-layer perceptron (MLP) algorithms produced the best results for all three IAQ sensors among all algorithms tested. The models built with data mining algorithms can serve as virtual IAQ sensors in buildings and be used for on-line monitoring and calibration of the IAQ sensors. The approach presented in this paper can be applied to HVAC systems in buildings beyond the type considered in this paper.

1. Introduction

Heating, ventilation, and air conditioning (HVAC) systems ensure good quality air inside buildings. Statistics show that maintaining the air comfort in buildings consumes more than 40% of the energy used in US commercial and residential buildings [1,2]. The growing complexity of building HVAC systems has become a major challenge for applying strategies for reducing costs and enhancing air quality [3]. Degraded equipment, failed sensors, improper installation, poor maintenance, and aged control systems have attributed to poor performance of HVAC systems of various commercial buildings.

Indoor air quality is usually measured by the level of temperature [°C], CO_2 [PPM], and relative humidity [%]. The performance of IAQ sensors greatly impacts the air quality and energy savings of HVAC systems. Various researchers discussed modeling sensors, detection of sensor faults, and developing cost-efficient control strategies for HVAC systems [4–9]. Namburu et al. [10] proposed a real-time fault diagnosis scheme for chillers based on a data-driven approach. Cho et al. [11] developed a model based on pattern classification and residual ratios to diagnose, identify, and detect multiple-faults occurring in HVAC systems. Salsbury and Diamond [12] used simulation to predict performance targets and compare monitored system outputs for performance validation and energy analysis. Hou et al. [13] combined a rough set approach with a neural network algorithm to build a model based on past HVAC performance data. The model was intended to detect and diagnose sensor faults in HVAC systems. Schein et al. [14] developed a rule-based method for fault detection in air handling systems. The rules were derived by experts from mass and energy balances.

Control charts are used in process monitoring to determine and eliminate the sources of process variation so that the process returns to its normal state. They have been widely researched in the statistical quality and process control literature [15]. Recent advances in profile monitoring have led to applications in manufacturing, calibration, logistics, service, marketing, finance, and accounting [16–20].

Identifying statistical control process profiles with high fidelity is needed in most process applications. However, the majority of literature has focused on parametric profile monitoring, in particular linear profiles. Developing non-linear non-parametric models characterizing indoor air quality is a challenge for IAQ sensors used in HVAC systems. Data mining offers promising approaches to modeling IAQ sensors. Numerous applications of data mining in manufacturing [21,22], marketing [23,24], medical informatics [24,25], and the energy industry [26,27] have proven to be effective in support of decision making and optimization. Some applications of data mining algorithms has been reported for HVAC systems [28,29].

In this paper, data mining algorithms have been applied to build models of IAQ sensors. The constructed models are used as on-line profiles and virtual sensors for indoor air quality. They can be also

* Corresponding author. Tel.: +1 319 3355934; fax: +1 319 3355669.
E-mail address: andrew-kusiak@uiowa.edu (A. Kusiak).

0306-2619/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apenergy.2009.12.008

used to monitor the performance of IAQ sensors installed in HVAC systems. The IAQ sensor models are built based on historical data from an HVAC system at an existing facility.

2. Data description and methodology for modeling and monitoring of IAQ sensors

2.1. HVAC data

The data used in this research was collected at the Iowa Association of Municipal Utilities (IAMU) Office Building and Training Complex, which included 12,500 sq ft of office space, maintenance, and shop facilities. The HVAC control system collected data for more than 60 parameters with a sampling interval of 1 min. Due to the type of data collection system available for this research, the values stored for all HVAC parameters are the (last-measured) point data rather than the 1-min average data. The HVAC data from the sensors installed at the IAMU was supplemented with the outside weather data collected by the Iowa Energy Center over a 2-month period.

For each IAMU office/room of interest to this research, the data was collected for the same set of IAQ parameters: temperature, CO_2, and relative humidity, as well as other parameters. In this paper, the IAMU auditorium was selected for in-depth analysis; however, the approach presented can be applied to any office in the building. Table 1 lists the HVAC parameters of the IAMU auditorium used in this research. The first three parameters in Table 1 are the indoor air quality parameters collected from the IAQ sensors; the parameter "Aud_IAQ_Temp" is the temperature measured at the thermostat installed in the auditorium; the last eight parameters indicate the outside weather conditions collected by the Iowa Energy Center.

The data set used for IAQ sensor modeling was divided into two independent data subsets, a training data set and a test data set. The training data set was used to develop models of IAQ sensors, while the test data set was used to validate the performance of the models learned from the training data set. Fig. 1a–c shows typical plots of temperature, CO_2 and relative humidity, respectively over a 1-day period. It can be observed that the IAQ parameter values change during the day. The horizontal axis refers to the data point number.

The values of the temperature, CO_2, and relative humidity are continuously changing. However, the IAQ sensor plots shown in Fig. 1 are not all continuously-changing curves, rather they contain steep changes. Various reasons could explain the abnormal spikes, including the degraded performance of sensors, faulty HVAC control, and unexpected working conditions around sensors. If accurate and robust virtual models of IAQ sensors were available, the physical IAQ sensors could be monitored and calibrated on-line or even replaced by the virtual IAQ sensors.

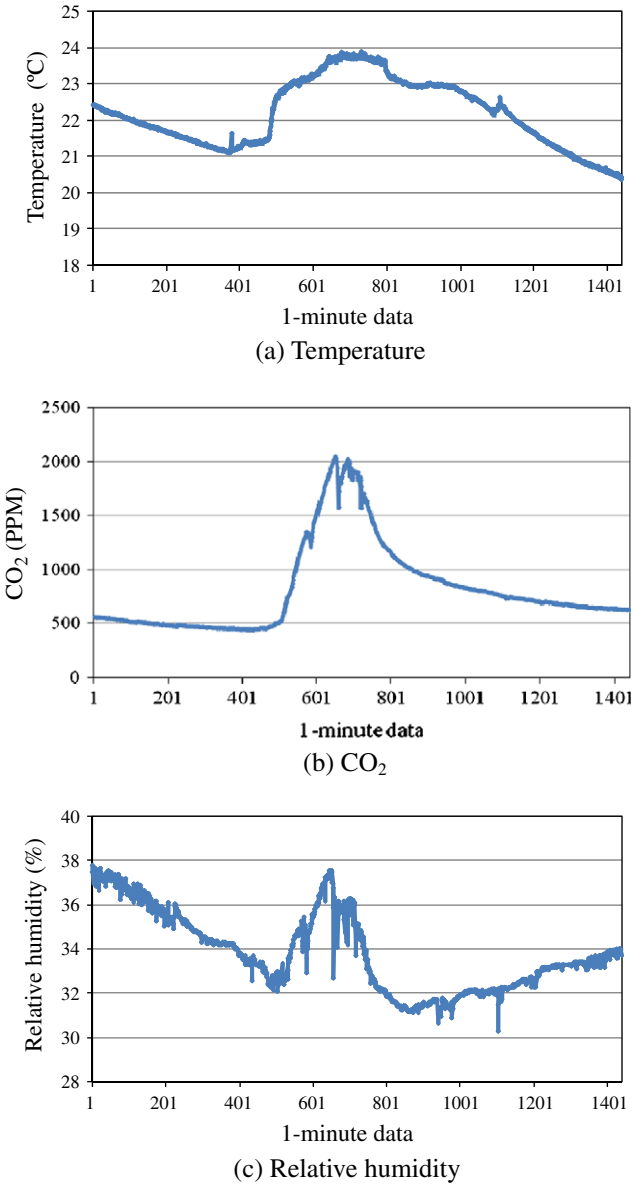

Fig. 1. Illustrative plots of IAQ parameters.

Table 1
Parameter description of the HVAC data set.

Parameter	Description	Unit
Aud_IAQ_Temp	Auditorium temperature from IAQ sensor	°C
Aud_IAQ_ CO_2	Auditorium CO_2 from IAQ sensor	PPM
Aud_IAQ_RH	Auditorium relative humidity from IAQ sensor	% RH
Aud_Temp	Controllable auditorium temperature	°C
Aud_Lite	Auditorium light level	FC
BAR-PRES	Barometric pressure	Bar
OA-HUMD	Relative humidity	% RH
OA-TEMP	Dry-bulb temperature	°C
SOL-BEAM	Direct normal solar irradiation	Btu/hr-ft^2
SOL-HORZ	Total horizontal irradiation	Btu/hr-ft^2
WIND-DIR	Wind direction	°
WIND-VEL	Wind speed	Mile/h

2.2. Virtual IAQ sensor modeling

Data mining algorithms were used to build models for the IAQ sensors in the HVAC system. Virtual IAQ sensor modeling used other HVAC parameters as predictors to predict IAQ parameters as dependent and the IAQ parameters included temperature, CO_2, and relative humidity. The relationship between IAQ parameters and various other HVAC parameters are complicated, and thus it is hard to identify the model and accurately predict IAQ parameters with high-dimension HVAC parameters as input using mathematical modeling. However, data mining is a powerful tool in extracting knowledge from voluminous data.

A virtual IAQ sensor model represents the underlying function between the IAQ parameter and the other HVAC parameters. Eqs. (1)–(3) show the sensor models for predicting temperature, CO_2, and relative humidity, respectively.

$$y_{\text{Aud\_IAQ\_Temp}} = f(x_{\text{Aud\_IAQ\_CO}_2}, x_{\text{Aud\_IAQ\_RH}}, v_{\text{Aud\_Temp}}, v_{\text{Aud\_Lite}} v_{\text{BAR-PRES}}, v_{\text{OA-HUMD}}, v_{\text{OA-TEMP}}, v_{\text{SOL-BEAM}}, v_{\text{SOL-HORZ}}, v_{\text{WIND-DIR}}, v_{\text{WIND-VEL}}) \quad (1)$$

$$y_{\text{Aud\_IAQ\_CO}_2} = f(x_{\text{Aud\_IAQ\_RH}}, x_{\text{Aud\_IAQ\_Temp}}, v_{\text{Aud\_Temp}}, v_{\text{Aud\_Lite}} v_{\text{BAR-PRES}}, v_{\text{OA-HUMD}}, v_{\text{OA-TEMP}}, v_{\text{SOL-BEAM}}, v_{\text{SOL-HORZ}}, v_{\text{WIND-DIR}}, v_{\text{WIND-VEL}}) \quad (2)$$

$$y_{\text{Aud\_IAQ\_RH}} = f(x_{\text{Aud\_IAQ\_CO}_2}, x_{\text{Aud\_IAQ\_Temp}}, v_{\text{Aud\_Temp}}, v_{\text{Aud\_Lite}} v_{\text{BAR-PRES}}, v_{\text{OA-HUMD}}, v_{\text{OA-TEMP}}, v_{\text{SOL-BEAM}}, v_{\text{SOL-HORZ}}, v_{\text{WIND-DIR}}, v_{\text{WIND-VEL}}) \quad (3)$$

In Eqs. (1)–(3) y is the dependent IAQ parameter, x the IAQ parameter used in this model as a predictor, and v is the parameter indicating the outside weather conditions. The v and x parameters are listed in Table 1 of Section 2.1. The model $f(\cdot)$ is learned by a data mining algorithm. An obvious advantage of the data-driven approach is that $f(\cdot)$ can be easily and timely updated by the most current HVAC process data. Deriving an accurate virtual IAQ model that maps complicated relationships among the parameters of the HVAC system is a challenge.

The selection of an appropriate data mining algorithm is important for building an accurate, stable, and robust IAQ model. Different data mining algorithms were applied for IAQ sensor modeling, and the performance of the various data mining algorithms was analyzed. Two basic metrics, the mean absolute error (MAE) and standard deviation of absolute error (Std) were used to compare prediction accuracy. They were computed to select the best data mining algorithm to extract the accurate IAQ sensor model (Eqs. (1)–(3)). The small value of the MAE and Std implies the superior prediction performance of the IAQ model. The absolute error (AE), mean absolute error (MAE), and the standard deviation (Std) are expressed in Eqs. (4)–(6).

$$\text{AE} = |\hat{y} - y| \quad (4)$$

$$\text{MAE} = \frac{\sum_{i=1}^{N} \text{AE}(i)}{N} \quad (5)$$

$$\text{Std} = \sqrt{\frac{\sum_{i=1}^{N} (\text{AE}(i) - \text{MAE})^2}{N-1}} \quad (6)$$

where $\hat{y}$ is the predicted IAQ parameter value, y the observed (measured by sensor) IAQ parameter value, and N is the number of test data points used to validate the performance of the IAQ sensor model learned by a data mining algorithm.

2.3. IAQ sensor monitoring based on control charts

The physical sensors installed in any HVAC system degrade over time, and this leads to inferior performance, poor air quality, and energy waste due to incorrect feedback from the degraded IAQ sensors. A formal approach for on-line monitoring of IAQ sensors is necessary. The IAQ sensor models built by data mining algorithms and control charts borrowed from statistical process control theory can be used to detect and remedy performance anomalies. Identifying sensor faults and on-line monitoring of the IAQ sensors is useful for optimizing the performance of HVAC systems.

Fig. 2 illustrates the basic concept of IAQ sensor modeling and on-line monitoring presented in this paper. A data mining algorithm is used to identify IAQ sensor models based on the historical HVAC process data. The model can be updated to reflect the process change over time. The update frequency could be, for example, 2 weeks. The operational update frequency depends on the HVAC system operational conditions and the accuracy requirements. Alternatively, a separate routine could monitor the model performance and refresh the model once its performance degraded. A control chart generated from the HVAC data can be used for on-line monitoring of an IAQ sensor. The IAQ models and control chart monitor the IAQ sensor performance at a certain time interval, e.g., 5 min.

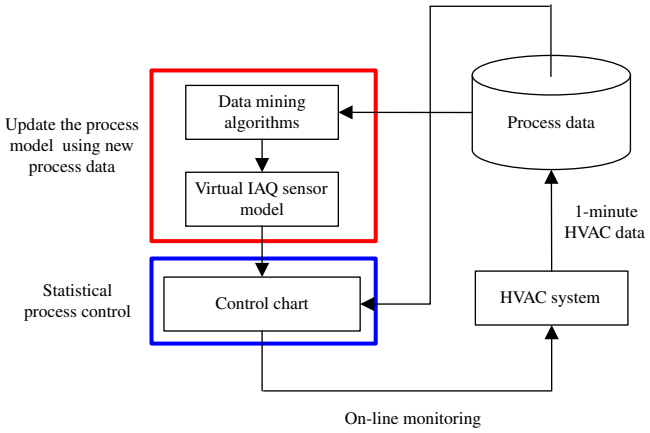

Fig. 2. Modeling and on-line monitoring of IAQ sensors.

Data mining algorithms identify models for IAQ sensors that can serve as on-line indoor air quality profiles for temperature, CO_2, and relative humidity, respectively. The residual control chart approach (statistical quality control) [17,18] is used to analyze residuals between the model predicted IAQ value and the observed (measured by sensor) IAQ value. The residual ε is expressed as [18]:

$$\varepsilon = \hat{y} - y \quad (7)$$

where y is the observed IAQ value and $\hat{y}$ is the reference value predicted by an IAQ sensor model.

The control chart approach [16–18] allows the residuals and their variations to be monitored, and thus detect abnormal conditions and an IAQ sensor fault. A training data set of N_{train} observations with outliers removed was selected to build a control chart. The training data set is represented as

$$y\_TrainSet = [y(i), \hat{y}(i)], \quad i = 1, \ldots, N_{train}.$$

Using the training data set, the residual ε for each point is computed, as well as the mean and the standard deviation of ε. The mean residual μ_{Train} and the standard deviation σ_{Train} are shown as [30]:

$$\mu_{Train} = \frac{1}{N_{train}} \sum_{i=1}^{N} (\hat{y}(i) - y(i))$$

$$\sigma_{Train} = \sqrt{\frac{1}{N_{train} - 1} \sum_{i=1}^{N} ((\hat{y}(i) - y(i)) - \mu_{Train})^2} \quad (8)$$

The test data set $y\_TestSet = [y(i), \hat{y}(i)]$ includes N_{test} consecutive data points drawn in time sequence from the test data set shown in Table 4.

$$\mu_{Test} = \frac{1}{N_{test}} \sum_{i=1}^{n} (\hat{y}(i) - y(i))$$

$$\sigma_{Test} = \sqrt{\frac{1}{N_{test} - 1} \sum_{i=1}^{n} ((\hat{y}(i) - y(i)) - \mu_{Test})^2} \quad (9)$$

Once μ_{Train} and σ_{Train} are known, the upper and lower control limits of the control chart are computed and used to detect anomalies. Based on Eq. (8), the control limits of the control chart are derived from [16,17]:

$$\begin{aligned} UCL_1 &= \mu_{Train} + \eta\frac{\sigma_{Train}}{\sqrt{N_{test}}} \\ CenterLine_1 &= \mu_{Train} \\ LCL_1 &= \mu_{Train} - \eta\frac{\sigma_{Train}}{\sqrt{N_{test}}} \end{aligned} \tag{10}$$

N_{test} is the number of points in *y_TestSet*, η is the integer multiple for the control limits, and N_{test} (usually fixed as 3) can be adjusted to make the control chart less sensitive to the data variability and thus reduce the risk of false alarms. In this paper, N_{test} was set at 5 to make the control chart less sensitive to the data variability. If μ_{Test} is above UCL_1 or below LCL_1, the IAQ parameter value at the sampling time *y_TestSet* is considered to be deficient, and this type of fault detected by the control chart is defined as Fault Type I.

Similarly, the control limits for σ^2_{Test} are calculated as [16,17]:

$$\begin{aligned} UCL_2 &= \frac{\sigma^2_{Train}}{N_{test}-1} \times \chi^2_{\alpha/2,N_{test}-1} \\ CenterLine_2 &= \sigma^2_{Train} \\ LCL_2 &= 0 \end{aligned} \tag{11}$$

where $\chi^2_{\alpha/2,N_{test}-1}$ denotes the right $\alpha/2$ percentage points of the chi-square distribution, $N_{test}-1$ is the degree of freedom of the chi-square distribution. The parameter α needs to be adjusted to make the control chart less sensitive to the variability of the data. LCL_2 is set to 0 to indicate that the variation of residuals in the test data is 0, so that the measured IAQ value matches the reference IAQ value in the normal status. If σ^2_{Test} is above UCL_2, the IAQ parameter value at the sampling time *y_TestSet* is considered as deficient, and this type of fault detected by control chart is defined as Fault Type II.

Table 2
The 2-day-long data set characterization.

Data set	Sample size (%)	Description
1	100	Total data set; 2880 observations
2	80	Training data set; 2304 observations
3	20	Test data set; 576 observations

3. Case study

To test the methodology presented in this paper, the HVAC data from the Iowa Association of Municipal Utilities (IAMU) Office Building and Training Complex in Ankeny, Iowa, was used.

3.1. Algorithm selection for IAQ sensor modeling

Different data mining algorithms have been used to extract the IAQ sensor models (Eq. (1)–(3)) from the HVAC data set. To select the best data mining algorithm for constructing accurate and robust IAQ sensor models, a 2-day-long data set was used. Table 2 shows data set 1 with a beginning time stamp of "3/30/2005 0:00" and an ending time stamp of "3/31/2005 23:50". Data set 1 was divided into two data subsets using a random sampling method, data sets 2 and 3. Data set 2 contains 2304 randomly selected data points and was used to develop a model $f(\cdot)$ estimating the IAQ parameters. Data set 3 includes 576 randomly selected data points and was used to test the performance of the model $f(\cdot)$ learned from data set 2.

Four different algorithms were used to learn (identify) the IAQ sensor models. They included the Multi-layer perceptron (MLP), neural network (NN) [25,31], Radial-basis-function (RBF) NN [31,32], support vector machine regression (SVM) [33,34], and Pace regression (Pacereg) [24,35]. The MLP NN and RBF NN algorithms are usually used in non-linear regression and classification modeling due to their ability to capture complex relationships between parameters. The BFGS (Broyden–Fletcher–Goldfarb–Shanno) algorithm minimizing the sum of square cost function is used in building the neural network. The SVM is a supervised learning algorithm used in classification and regression. It constructs a linear discriminant function separating instances (examples) as widely as possible. For ease of computation, in the SVM regression a high dimensional linear optimization problem is transformed into a dual convex quadratic programming problem. The Pacereg algorithm uses a group of estimators that are either overall optimal or optimal under certain conditions. It is a relatively new approach for developing linear models in high dimensional spaces.

In this research 35 MLP NN and 35 RBF NN models with different kernels and structures were built, and the most accurate and robust model was selected based on the objective function ex-

Table 3
Prediction accuracy of different algorithms.

Algorithm	Parameter	MAE	Std	Max	Min
MLP	Temp (°)	0.0311	0.0276	0.3052	0.0000
	CO_2 (PPM)	6.4238	8.5631	86.1721	0.0133
	RH (%)	0.1111	0.1096	1.2297	0.0004
RBF	Temp (°)	0.9478	0.6593	2.9145	0.0096
	CO_2 (PPM)	184.1855	208.7004	799.9020	0.1212
	RH (%)	1.7544	0.9513	4.7719	0.0069
Pacereg	Temp (°)	0.1668	0.1474	0.7421	0.0007
	CO_2 (PPM)	71.3895	78.2013	534.2520	0.3677
	RH (%)	0.5458	0.4051	2.7928	0.0019
SVM	Temp (°)	0.1473	0.1951	0.9691	0.0000
	CO_2 (PPM)	61.1816	106.5812	575.9267	0.0069
	RH (%)	0.5343	0.4859	3.3849	0.0015

Table 4
The description of 2-week data set.

Data set	Start time	End time	Description
1	05/01/2007 00:00	05/11/2007 23:50	Total data set; 15,840 observations
2	05/01/2007 00:00	05/04/2007 23:50	Training data set; 5760 observations
3	05/07/2007 00:00	05/09/2007 23:50	Training data set; 4320 observations
4	05/10/2007 00:00	05/11/2007 23:50	Test data set; 2880 observations

pressed as the sum of square errors between observed and predicted values. Five different activation functions were used for the hidden and output neurons, namely, the logistic, identity, tanh, exponential, and sine function. The number of hidden units was set between 4 and 14, and the weight decay for both hidden and output layer varied from 0.0001 to 0.001.

Table 3 summarizes the prediction performance of IAQ sensor models learned by different data mining algorithms. The performance of data mining algorithms is ranked using the MAE (mean absolute value) of Eq. (5)) and Std (standard deviation) of Eq. (6) for the test data set from Table 2. The small value of the MAE and Std indicates accurate, stable, and robust prediction performance. The MLP algorithm performed best among the four data mining algorithms due to its smallest MAE and Std in all three IAQ sensor models, while the RBF performed the worst, as indicated by the largest value of MAE and Std. The Pacereg and SVM algorithms performed quite well for prediction of temperature and relative humidity. The MLP algorithm provided high quality predictions for the test data set and captured the HVAC system dynamics with high fidelity. The performance of models built by the MLP algorithm is validated in Section 3.2, based on a larger data set (see Fig. 3).

Fig. 4a–c shows respectively the first 100 predicted (by MLP) and observed temperature, CO_2, and relative humidity computed for the test data set of Table 2. It is easy to see that the observed and predicted values of IAQ parameters are almost identical and even overlap, and the predicted value follows exactly the trend of the observed value. The three IAQ sensor models built by MLP algorithm predict accurately CO_2, temperature and relative humidity, respectively, on the data set 3 of Table 2.

3.2. Validation of IAQ models based on a 2-week data set

The MLP NN algorithm performed best among the four different data mining algorithms based on the 2-day data set, and thus it was selected for further investigation on a 2-week data set. Table 4 shows the 2-week data set with a beginning time stamp of "5/1/2007 0:00" and an ending time stamp of "5/11/2007 23:50". Data set 1 was divided into three data subsets, data sets 2–4. All the data

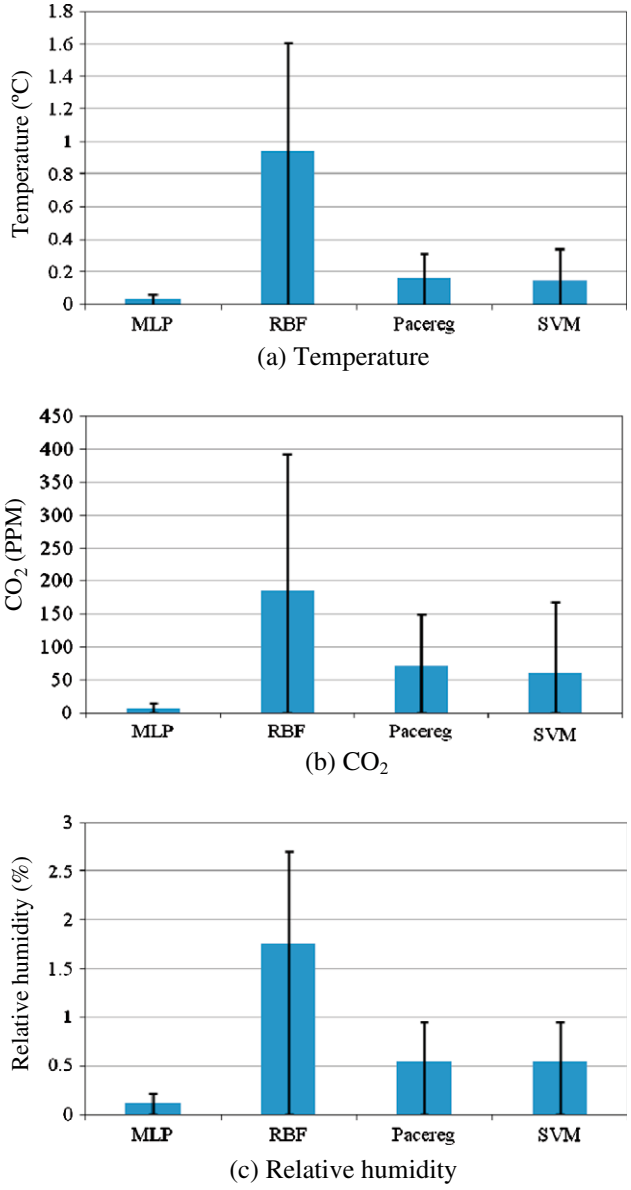

Fig. 3. The bar-chart of prediction performance of different algorithms for test data set 3 of Table 2.

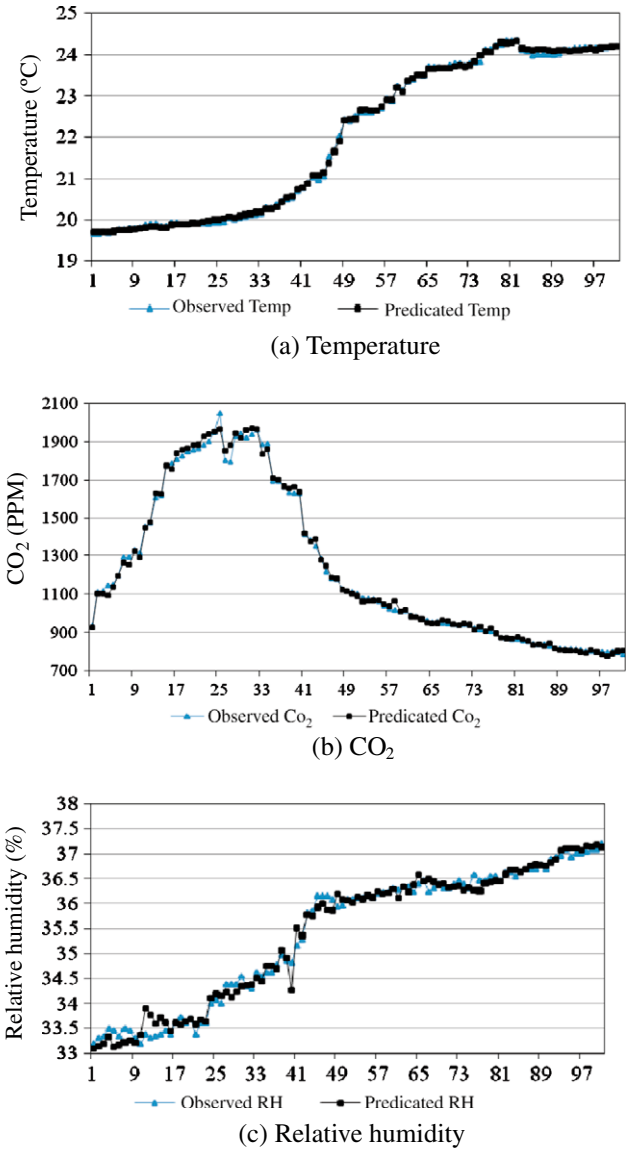

Fig. 4. Predicted and observed value of IAQ parameters of the test data set from Table 2.

sets contain consecutive data points drawn in a time sequence. Considering that the HVAC system of the IAMU building operates at different conditions during weekdays and weekends, the weekend data of 5/5/2007 and 5/6/2007 was removed from data set 1, and thus two training data sets were created due to a weekend (2-day) data set between the 2 weekday data sets. Data sets 2, containing 5760 time-consecutive data points, and data set 3, containing 4320 time-consecutive data points, were used to develop a model $f(\cdot)$ estimating the IAQ parameters. Data set 4, containing 2880 time-consecutive data points, was used to test the performance of the models $f(\cdot)$ learned from data set 2 and data set 3.

Fig. 5a–c shows the first 200 predicted (by MLP) and observed temperature, CO_2 and relative humidity for the test data set of Table 4. The observed and predicted values of IAQ parameters match well, and the predicted values follow exactly the trend of the observed value or even overlap. On the 2-day data set of Table 4, the three IAQ models constructed by MLP algorithm perform satisfactorily. The horizontal axis denotes the data point sequential number.

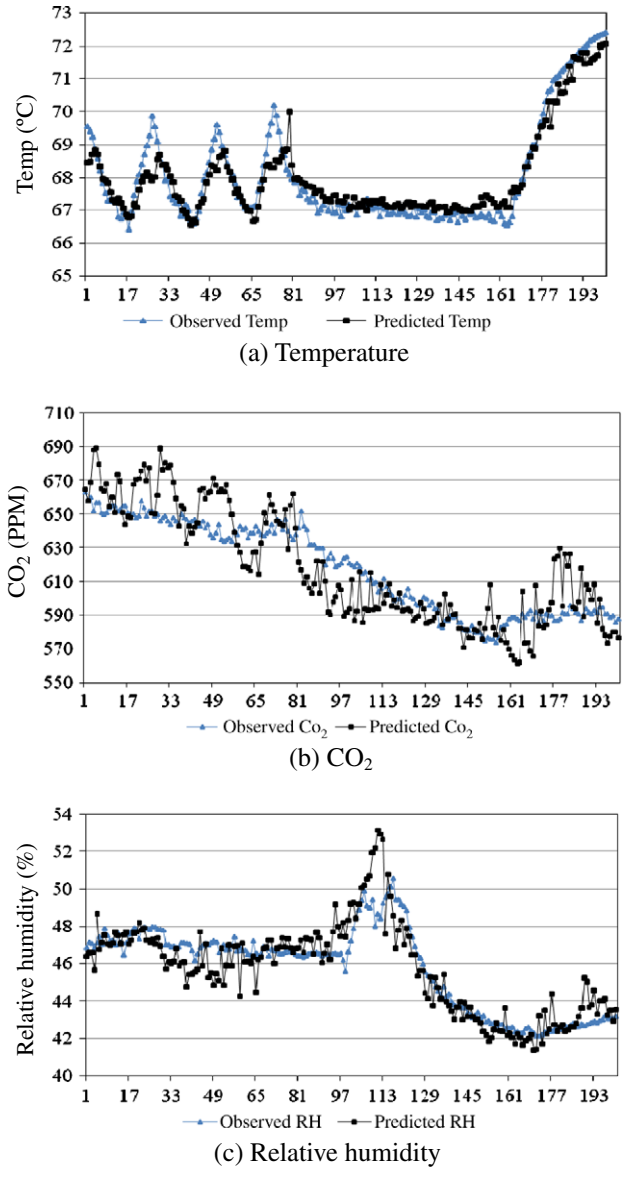

Fig. 5. Predicted and observed values of IAQ parameters for the data set 4 from Table 4.

Table 5
Prediction accuracy of different algorithms for the test data set of Table 4.

Parameter	Mean	Std	Max	Min
Temp (°)	0.2749	0.2815	2.1389	0.0001
CO_2 (PPM)	25.4649	28.9067	277.6084	0.0010
RH (%)	0.7443	0.6962	4.8757	0

Table 5 summarizes the prediction accuracy of three IAQ models built by the MLP algorithm. The MLP algorithm performs well on both the 2-day data set (data sets 2 and 3 of Table 4) and the 2-week data set (data set 4 from Table 4). The performance is maintained accurate and robust. However, in actual applications, updating the learned model with the new data is necessary for processes that are temporal. The temporal process modeling task is accomplished by using data mining algorithms. A large prediction error (MAE and Std) indicates that the model built on historical data needs to be updated. The update time is impacted by various factors, e.g., if the seasonal and rapid change of weather conditions.

The sensitivity of a neural network's output to its input perturbation is an important issue in theory and practice. The sensitivity analysis [31,32] indicates input variables that are most important for a particular neural network. It often identifies variables that can be safely ignored in subsequent analyses and key variables that must always be retained. Sensitivity analysis ranks variables according to the deterioration of the model performance if that variable is no longer available for the model. In so doing, it assigns a single rank value to each variable. Table 6 shows the sensitivity analysis of the three IAQ sensor models built by the MLP based on the 2-week data set. Different parameters have different sensitivity ranks in the three IAQ sensor models. The sensitivity analysis offers insight into the complicated and non-linear relationship between various HVAC parameters and the outside weather condition parameters.

3.3. On-line IAQ sensor monitoring with virtual sensor models

The approach and equations of Section 2.3 have been implemented to form control charts for the three IAQ sensor models. The three IAQ sensor models (Eq. (1)–(3)) are constructed by the MLP NN algorithm based on data sets 2 and 3 of Table 4, and the parameters of three control charts for CO_2, temperature and relative humidity are computed based on the same training data set. For the temperature and the relative humidity models, the number of hidden units is 14. The hidden unit activation function is hyperbolic, while the output unit activation function is logistic. For the CO_2 model, the number of hidden units is 11. The hidden unit activation function is logistic and the output unit activation function is exponential. The on-line simulation and validation of the three virtual IAQ sensor models are performed using the test data set of Table 4.

Table 7 shows the control chart parameters of the three IAQ sensor models formed based on the training data set of Table 4. The parameter in Eq. (11) is fixed as 2, and η of Eq. (10) is set to 3 to enhance the confidence of detecting anomalies in the IAQ sensors. Two important parameters α and η can be adjusted according to the real situation in practice to reduce the risk of a false alarm by considering the normal state as an abnormal one. A control chart monitors the residual mean and variation of the sampling data at the same time. In this research, the UCL_1 and LCL_1 calculated in Eq. (10) monitor the residual mean of the test data set, and the sensor fault detected due to the abnormal residual mean is defined as Fault Type I; $UCL2$ calculated from Eq. (11) monitors the residual variation of the test data set, and the sensor fault detected due to the abnormal residual variation is defined as Fault Type II.

Table 6
Sensitivity analysis of MLP NN for test data set of Table 4.

Temperature		CO_2		Relative Humidity	
Parameter	Sensitivity	Parameter	Sensitivity	Parameter	Sensitivity
Aud_Temp	797.5280	Aud_IAQ_Temp	207.6935	Aud_IAQ_Temp	305.3133
Aud_IAQ_ CO_2	235.3234	Aud_Temp	183.0385	Aud_IAQ_ CO_2	236.4275
BAR-PRES	94.3404	BAR-PRES	181.5638	BAR-PRES	188.0380
Aud_IAQ_RH	66.1309	Aud_IAQ_RH	101.6903	Aud_Temp	73.9207
OA-TEMP	34.5833	SOL-HORZ	66.4172	OA-TEMP	22.9534
OA-HUMD	10.6605	SOL-BEAM	54.0379	SOL-BEAM	15.5511
SOL-BEAM	6.3443	OA-TEMP	8.7827	SOL-HORZ	9.5619
SOL-HORZ	5.2809	Aud_Lite	3.2909	Aud_Lite	7.0401
Aud_Lite	2.8174	OA-HUMD	1.4673	OA-HUMD	5.7857
WIND-DIR	1.3425	WIND-DIR	1.1229	WIND-DIR	1.2634
WIND-VEL	1.0893	WIND-VEL	1.0449	WIND-VEL	1.0526

Table 7
Control limit values for IAQ parameters.

Parameter	UCL1	LCL1	UCL2	LCL2
Temp (°C)	0.2794	−0.2780	0.2330	0.0000
CO_2 (PPM)	51.6453	−49.6820	4278.0080	0.0000
RH (%)	1.1661	−1.1550	2.2448	0.0000

Fig. 6a–c shows the on-line monitoring simulation results of the control limits of the three IAQ sensor models trained by the MLP NN algorithm based on the training data set of Table 4; the observed IAQ parameter curves and the curve of the upper and lower limits were constructed from the test data set of Table 4; the first three hundred data points among the 2800 test data points of Table 4 are shown in three figures. The upper limits (boundary) of three normal IAQ parameters curves are calculated based on the UCL_1 of Eq. (10). The lower limits (boundary) of three normal IAQ parameter charts are calculated based on the LCL_1 of Eq. (10). The horizontal axis is the data point number.

Table 8 shows the statistics of the on-line monitoring simulation results of three IAQ sensor control charts based on data set 4 of Table 4. Fault Type I is the abnormal data point detected by the residual mean (Eq. (10)), and Fault Type II is the sensor fault detected by residual variation. Table 8 shows the number of Type I faults and Type II faults and their percentage among the total test data set (2800 data points in total) of Table 4. The RH control chart detected more sensor faults than CO_2 and temperature due to the fact that the control chart could be sensitive, the relative humidity could be hard to measure by sensors, the sensor could have drifted or operated in abnormal working conditions, or the prediction of IAQ models may not have been accurate and robust enough. As shown in Table 8, less faults of Type II has been detected. Therefore, the parameters in Eqs. (10) and (11) could make the three control charts more robust and less sensitive. These parameters could be adjusted dynamically based on the practical operations of an individual HVAC system.

4. Conclusion

In this paper, sensor models for predicting temperature, CO_2, and relative humidity were constructed by data mining algorithms. Four data mining algorithms were used to construct the IAQ sensor models. The MLP algorithm outperformed other data mining algorithms. A comprehensive comparative analysis of the IAQ sensor models built with different data mining algorithms and a sensitivity analysis of the three IAQ sensor models built by the MLP was reported in this paper. The performance of the selected MLP algorithm was validated on a 2-week data set. Data mining algorithms identified IAQ sensor models from the actual HVAC process data, and capture the complicated relationship between HVAC parameters and outside weather condition parameters.

The IAQ sensor models were used as the reference IAQ sensors (on-line profile) for monitoring the performance of physical IAQ

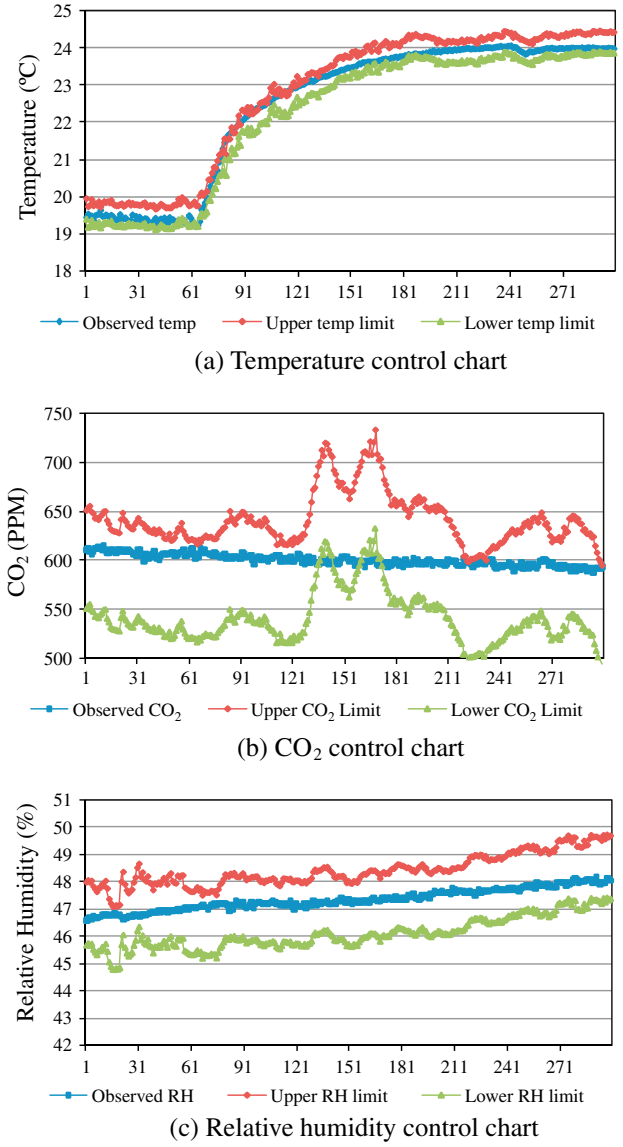

Fig. 6. Control charts of IAQ parameters with upper and lower limits for test data set of Table 4.

Table 8
The statistics of IAQ sensor fault detection of control charts.

IAQ parameter	Fault Type I (%)	Fault Type II (%)	Fault Type I	Fault Type II
Temperature (°)	9.4097	2.7431	271	79
CO_2 (PPM)	10.7292	1.8750	296	54
RH (%)	20.3125	0.5903	585	17

sensors and the indoor air quality of the HVAC system. The control chart approach can be used to monitor the residuals between the observed and the reference IAQ parameter values, and variation of the residuals. The IAQ sensor faults in non-stationary HVAC processes with certain structures can be detected in an on-line fashion. On-line IAQ sensor monitoring provides a data-driven approach for IAQ sensor fault detection for the HVAC industry. Three IAQ sensor models could be used as virtual sensors for monitoring, calibration, and even replacement of the physical sensor installed in HVAC systems. A virtual sensor estimates the value of a parameter in addition or in the absence of the actual physical sensor. The approach developed in this paper provides a basis for HVAC control optimization to sustain good indoor air quality and energy savings.

Constructing data-driven models with data mining algorithms is applicable to a wide range of processes in other domains. Additional data mining algorithms and concepts could be considered for enhancing the accuracy of the three IAQ sensor models. Control charts were able to detect the HVAC IAQ sensor faults, but could not identify the specific reasons or factors causing them. Further research is needed to implement the on-line fault detection and diagnosis system in HVAC systems.

Acknowledgement

This research has been supported by the Iowa Energy Center, Grant No. 08-01.

References

[1] Huang W, Zaheeruddin M, Cho SH. Dynamic simulation of energy management control functions for HVAC systems in buildings. Energy Convers Manage 2006;47(7–8):926–43.
[2] <http://www.fmlink.com/Marketplace/WhitePapers/Articles/kimberly-clark-040507.html>; 2008 [Accessed 19.09.08).
[3] Mathews EH, Botha CP, Arndt DC, Malan A. HVAC control strategies to enhance comfort and minimize energy usage. Energy Build 2001;33(8):853–63.
[4] Aktacir MA, Büyükalaca O, Yilmaz T. A case study for influence of building thermal insulation on cooling load and air-conditioning system in the hot and humid regions. Appl Energy 2010;87(2):599–607.
[5] Krüger EL, Laroca C. Thermal performance evaluation of a low-cost housing prototype made with plywood panels in Southern Brazil. Appl Energy 2010;87(2):661–72.
[6] Peeters L, de Dear R, Hensen J, D'haeseleer W. Thermal comfort in residential buildings: comfort values and scales for building energy simulation. Appl Energy 2009;86(5):772–80.
[7] Liang X, Chan MY, Deng S. Development of a method for calculating steady-state equipment sensible heat ratio of direct expansion air conditioning units. Appl Energy 2008;85(12):1198–207.
[8] Wong LT, Mui KW. A transient ventilation demand model for air-conditioned offices. Appl Energy 2008;85(7):545–54.
[9] Mui KW, Wong LT, Law LY. An energy benchmarking model for ventilation systems of air-conditioned offices in subtropical climates. Appl Energy 2007;84(1):89–98.
[10] Namburu SM, Azam MS, Luo J, Choi K, Pattipati KR. Data-driven modeling, fault diagnosis and optimal sensor selection for HVAC chillers. IEEE Trans Automat Sci Eng 2007;4(3):469–73.
[11] Cho S, Hong Y, Kim W, Zaheeruddin M. Multi-fault detection and diagnosis of HVAC systems: an experimental study. Int J Energy Res 2005;29(6):471–83.
[12] Salsbury T, Diamond R. Performance validation and energy analysis of HVAC systems using simulation. Energy Build 2000;32(1):5–17.
[13] Hou Z, Lian Z, Yao Y, Yuan X. Data mining based sensor fault diagnosis and validation for building air conditioning system. Energy Convers Manage 2006;47(15–16):2479–90.
[14] Schein J, Bushby ST, Castro NS, House JM. A rule-based fault detection method for air handling units. Energy Build 2006;38(12):1485–92.
[15] Woodall WH, Spitzner DJ, Montgomery DC, Gupta S. Using control charts to monitor process and product quality profile. J Qual Technol 2004;36(3):309–20.
[16] Mitra A. Fundamentals of quality control and improvement. Upper Saddle River (NJ): Prentice Hall; 1998.
[17] Montgomery DC. Introduction to statistical quality control. New York: John Wiley; 2005.
[18] Kang L, Albin SL. On-line monitoring when the process yields a linear profile. J Qual Technol 2000;32(4):418–26.
[19] Mestek O, Pavlik J, Suchanek M. Multivariate control chart: control charts for calibration curves. J Anal Chem 1994;350(6):344–51.
[20] Wheeler DJ. Understanding variation: the key to managing chaos. Knoxville (TN): SPC Press; 2000.
[21] Kusiak A, Mining Data. Manufacturing and service applications. Int J Prod Res 2006;44(18–19):4175–91.
[22] Backus P, Janakiram M, Mowzoon S, Runger GC, Bhargava A. Factory cycle-time prediction with data-mining approach. IEEE Trans Semiconduct Manuf 2006;19(2):252–8.
[23] Berry MJA, Linoff GS. Data mining techniques: for marketing, sales, and customer relationship management. New York: Wiley; 2004.
[24] Tan PN, Steinbach M, Kumar V. Introduction to data mining. Boston (MA): Pearson Education/Addison Wesley; 2006.
[25] Seidel P, Seidel A, Herbarth O. Multilayer perceptron tumor diagnosis based on chromatography analysis of urinary nucleoside. Neural Networks 2007;20(5):646–51.
[26] Kusiak A, Song Z. Combustion efficiency optimization and virtual testing: a data-mining approach. IEEE Trans Indus Informat 2006;2(3):176–84.
[27] Kusiak A, Zheng HY, Song Z. Models for monitoring of wind farm power. Renewable Energy 2009;34(3):583–90.
[28] Kusiak A, Li M-Y. Cooling output optimization of an air handling unit. Appl Energy 2010;87(3):901–9.
[29] Du Z, Jin X, Yang Y. Fault diagnosis for temperature, flow rate and pressure sensors in VAV systems using wavelet neural network. Appl Energy 2009;86(9):1624–31.
[30] Casella G, Berger R. Statistical inference. Pacific Grove (CA): Brooks/Cole; 1990.
[31] Bishop CM. Neural networks for pattern recognition. New York: Oxford University Press; 1995.
[32] Espinosa J, Vandewalle J, Wertz V. Fuzzy logic, identification and predictive control. London (UK): Springer-Verlag; 2005.
[33] Smola AJ, Schoelkopf B. A tutorial on support vector regression. Stat Comput 2004;14(3):199–222.
[34] Shevade SK, Keerthi SS, Bhattacharyya C, Murthy KRK. Improvements to the SMO algorithm for SVM regression. IEEE Trans Neural Networks 2000;11:1188–93.
[35] Witten IH, Frank E. Data mining: practical machine learning tools and techniques. San Francisco (CA): Morgan Kaufmann; 2005.