Indoor Air Quality Monitoring System for Smart Buildings

Xuxu Chen[1,2*], **Yu Zheng**[2,+], **Yubiao Chen**[2,3], **Qiwei Jin**[2], **Weiwei Sun**[1], **Eric Chang**[2], **Wei-Ying Ma**[2]
[1]Fudan University, Shanghai, China
[2]Microsoft Research, Beijing, China
[3]Harbin Institute of Technology, Harbin, China
{v-xuche, yuzheng, v-yubche, qiwjin, echang, wyma}@microsoft.com, wwsun@fudan.edu.cn

ABSTRACT
Many developing countries are suffering from air pollution, especially the Particulate Matter with diameter of 2.5 micrometers or less (PM2.5). While quite a few air quality monitoring stations have been built by governments in a city's public areas, the indoor PM2.5 has not yet been monitored and dealt with effectively. Though many office buildings have an HVAC (heating, ventilation, and air conditioning) system, PM2.5 is not considered as a factor when the system circulates fresh air from outdoors. This paper introduces a real system that we have deployed in the offices of four Microsoft campuses in China. This system instantly monitors indoor air quality on different floors of a building (including office areas, gyms, garages, and restaurants), enabling Microsoft employees to enquire the air quality of a place by using a mobile phone or checking a website. The information can guide a user's decision making, e.g., finding the right time to work out in the gym or turn on individual air filters in her own office. Through analyzing the indoor and outdoor air quality data collected over a long period, our system can even offer actionable and energy-efficient suggestion to HVAC systems, e.g., automatically turning on the system only a few hours earlier than usual if it is a heavily polluted day, or identifying the filters in HVAC system that should be renewed.

Author Keywords
Urban computing; smart building; Air quality; PM2.5

ACM Classification Keywords
H.2.8 [**Database Management**]: Database Applications - *data mining, Spatial databases and GIS*;

General Terms
Algorithms, Experimentation.

INTRODUCTION
Many developing countries, such as China, India, Mexico, and Brazil, are struggling with air pollution, especially PM2.5. To protect people's health from the damage by air pollution, many cities have built on-ground air quality monitoring stations that inform people the concentration of air pollutants in (outdoor) public areas [10]. While people stay indoors much longer than outdoors, the indoor PM2.5 has not yet been monitored effectively. As a result, people working in offices have no idea about the air quality around them, let alone taking actions to tackle PM2.5 down indoors.

In contrast to outdoor air pollutions that are difficult to tackle [3][12], the indoor PM2.5 can be handled to some extent if we manipulate HVAC systems or individual air filtering systems timely and correctly. Unfortunately, PM2.5 is not considered as a factor when HVAC systems circulate fresh air from outdoors. Some research projects [5][11] monitor the indoor concentration of CO_2; however, they do not provide actionable suggestions that can handle air quality problems. Additionally, sensing CO_2 is different from PM2.5, which needs a bigger sensor and a longer sensing period.

To address this issue, we deployed a cloud-based indoor air quality monitoring system in the office buildings of four Microsoft campuses in China [16] (consisting of Beijing, Shanghai, Wuxi, and Suzhou), as illustrated in Figure 1.

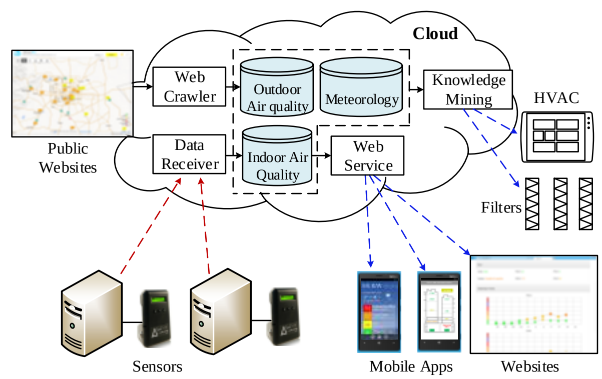


Figure 1. The architecture of our System

We collect the concentration of PM2.5 and PM10 on different floors of a building, including office areas, gyms, garages, and restaurants, etc. On a floor, we set up a monitor (Dylos DC1700) which is connected to a local server via a Com-to-USB port. The server receives the air quality readings from the monitor every minute and submits an average of air quality in every 10 minutes to the cloud. The cloud stores the air quality data received from different monitors in a cloud database, which will be enquired by end

UbiComp '14, September 13 - 17, 2014, Seattle, WA, USA
Copyright 2014 ACM 978-1-4503-2968-2/14/09…$15.00.
http://dx.doi.org/10.1145/2632048.2632103

**The research was done when the first and third authors were interns in Microsoft Research under the supervision of the second author +.*
+ Yu Zheng is the correspondence author of this paper.

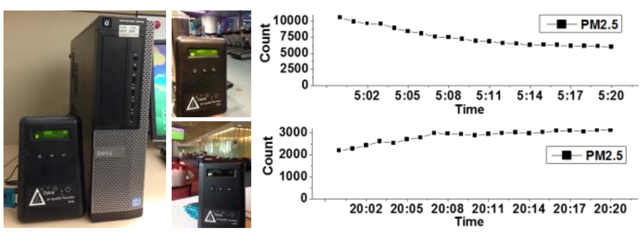
users through a mobile client and a website. The real-time air quality information can inform a user's decision making on when to work out in a gym or whether turning on an additional air filter in her own office.

The cloud also collects the outdoor air quality of each building and corresponding meteorological data from public websites every hour. The information will be displayed on the mobile client and website together with the indoor air quality. By mining the air quality and meteorological data over a long period, we build a model based on artificial neural network to suggest the number of hours that an HVAC system should be turned on ahead of its original schedule. The model considers the current outdoor and indoor air quality as well as meteorological data to make an inference. The model can also identify the floor where HVAC no longer works well, indicating that the air filter sheets of this floor should be replaced.

INDOOR PM2.5 MONITORING

Sensing

To detect the indoor concentration of PM2.5, we deploy an aerosol particle counter (Dylos DC1700) on each floor, as demonstrated in Figure 2 A). The particle counter measures the number of particles with a size bigger than $0.5\mu m$ but smaller than $2.5\mu m$ in each cube centimeter by using X-ray laser. The particle counter is connected to a local server via an USB-to-Com port adapter, streaming out the number of particles every minute, as illustrated in Figure 2 B). The local server then converts the received number into a concentration of PM2.5 ($\mu g/m^3$) through an empirical formula and submits the average concentration of every 10 minutes to the cloud.



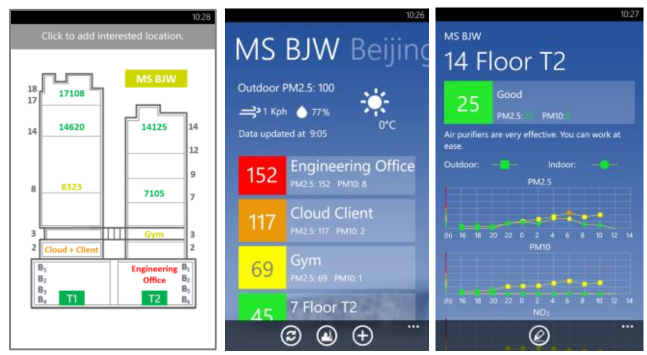
A) Sensors **B) Number of particles in one day**

Figure 2. Mobile User Interface

On the cloud, we run a web crawler to collect the outdoor concentration of PM2.5 and meteorological data, consisting of humidity, wind speed, temperature, and barometer pressure, from public websites every hour. The information is used to measure the effectiveness of an HVAC system in filtering the PM2.5 absorbed from outside and also employed as features in our model to infer the number of hours to turn on an HVAC ahead of its original schedule.

Displaying

Figure 3 A) visualizes the 2D map of Microsoft campus in Beijing, where 8 sensors have been deployed on different floors of the two towers (four sensors in each tower). The figure on each floor represents the location ID of a deployed sensor, with a color representing its AQI (Air Quality Index) level, e.g. "*green*" means "*good*" and "yellow" denotes "moderate" in Chinese AQI standard [14]. The color of block "T1" stands for the average AQI reported by the four sensors deployed in Tower 1. So does the color of "T2". The color of block "MS BJW" shows the average AQI level reported by all the 8 sensors in the two towers. Additionally, users can add the location they are concerned with into a location list demonstrated in Figure 3 B), by clicking on the floor shown on the 2D map. In Figure 3 B), each banner represents a location, e.g. *Engineering Office* and *Gym*. The two numbers associated with each banner denote the AQIs of PM2.5 and PM10, respectively. The color of a banner is determined according to its AQI levels. The outdoor weather information is also exhibited at the top. After clicking a specific banner, a user can check the trend of indoor and outdoor air quality, as shown in Figure 3 C). The effectiveness of the HVAC in filtering PM2.5 (or PM10) can be evaluated through the gap between outdoor and indoor AQIs at the same timestamp. In order to facilitate PC users, we also deploy a website showing same information available on the mobile client.



A) Select a location **B) Location list** **C) Trend of air quality**

Figure 3. Mobile User Interface

SMART SUGGESTION TO HVAC

Energy-Efficient Control on HVAC

In recent years, buildings have become one of the major energy consumers which account for almost 40 percent of energy consumption in the whole society [9]. The HVAC system as one of the major energy consumers in a building is usually turned off (or partially turned down) in the evening and turned on in the morning shortly before people start working in the building. In order to provide a healthy working environment to employees while saving energy, we predict the purification time (PT), i.e. a time period needed for an HVAC system to reduce indoor PM2.5 to an ideal situation, and turn on the HVAC at least PT hours before people's arrival. Figure 4 shows the definition of the PT in two scenarios. In scenario 1, the PT is defined as the time period $(t_1 - t_0)$ to reduce the concentration of PM2.5 to below $35\mu g/m^3$, which is regarded as "*good*" in Chinese AQI standard. In scenario 2, an HVAC system cannot reduce the indoor PM2.5 concentration below that threshold, given a certain high concentration of outdoor air quality and the limitation of the HVAC system. In this scenario, we regard the start point of a stable period like t_2 (i.e. the indoor PM2.5 concentration no longer decreases in the following 30 minutes) as the ending of the purification time. In the example shown in Figure 4, the PT is $t_1 - t_0$= 40 minutes in scenario 1 and $t_2 - t_0$= 60 minutes in scenario 2.

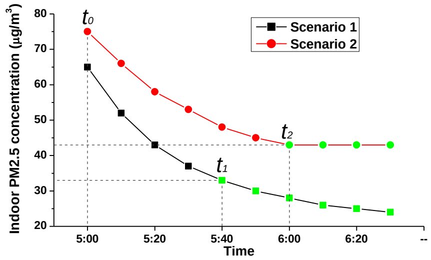


Figure 4. Purification time

An example: Suppose the majority of people start working in a building at 8am. The original schedule of turning on the HVAC system is 7am. There is a day with the concentration of outdoor PM2.5 much higher than usual. According to the prediction, the HVAC could need 1.5 hours to reduce the concentration to under $35\mu g/m^3$. To provide people with a healthy working environment on their arrivals, we need to turn on the HVAC at 6:30am, half hour earlier than the original schedule. Note that we do not change the operating strategy of an HVAC system, which considers multiple factors, such as the concentration of CO_2 and O_2. Turning on an HVAC system a few minutes ahead of its schedule is a safe action that does not break other environmental criteria.

Features

By analyzing the data (12/23/2013-5/9/2013), we notice that the purification time is influenced by multiple factors, such as the indoor and outdoor air quality, humidity, and barometer pressure, as illustrated in Figure 5, where each row and column denote one factor. Each plot in the figure stands for a PT we observed from the historical data, and different symbols represent different lengths of PT, e.g. a circle means 40-80minuts. For instance, the vertical axis of the box standing in the third row and the fifth column denotes outdoor humidity and its horizontal axis represents outdoor wind speed. It can be observed that high humidity and low wind speed cause a long purification time.

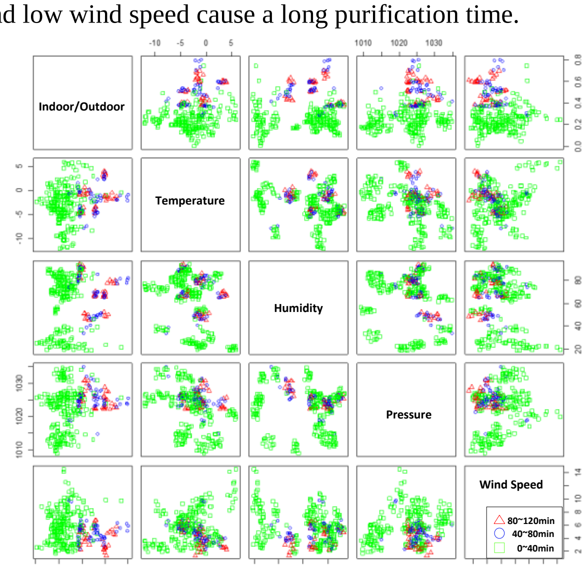


Figure 5. Correlation between purification time and features

Purification Time Inference (PTI)

We propose a Purification Time Inference (PTI) model based on artificial neural network (ANN), as illustrated in Figure 6. Specifically, PTI model is a three-layer network, with six nodes on the input layer, 16 nodes on the hidden layer, and 12 nodes on the output layer. Each node on the output layer denotes a certain length of PT, ranging from 10 –120 minutes (the maximum PT is 120 minutes in the historical data). The output value for the *i*th node is the probability that PT is i ×10 minutes. We then choose the most likely purification time C among the 12 values as our final result, which is defined as Equation 1:

$$C = \max_{1\le k\le 12} (\varphi(\textstyle\sum_{j=1}^{16} w'_{jk}\varphi(\sum_{i=1}^{6} w_{ij}F_i + b_j) + b'_k)), \quad (1)$$

where φ is a sigmoid function; F_i is the *i*th feature; b_j and b'_k are the biases associated with the nodes in hidden layer and output layer respectively; w_{ij} is the weight between input layer and hidden layer while w'_{jk} denotes the weight between hidden layer and output layer. All the parameters are trained with a Back-Propagation algorithm. The system performs the PTI model every 10 minutes, and notifies a building's operation team if the gap between the current time and people's arrival time is close to the inferred PT.

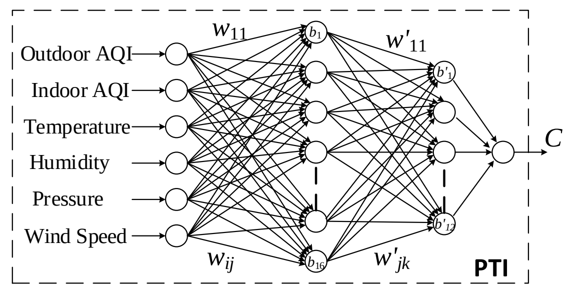


Figure 6. Framework of PTI

Renew HVAC's Air Filter Sheets

The inferred PT can also be used to identify the floor where the HVAC no longer works well, which could trigger an inspection on the floor's filter sheets. The assumption is that the real PT should be close to the inference in a normal situation. Specifically, if the real PT of a particular floor is longer than the inference by a threshold in consecutive days, our system sends an alarm. Figure 7 shows the real and inferred PTs of a floor in Beijing campus from 1/10/2014 to 3/10/2014. There was a significant gap between the real and inferred PTs around 2/21/2014. An inspection on the floor's HVAC found the filter sheets were very dirty and needed to be replaced. After the replacement, the gap is disappeared.

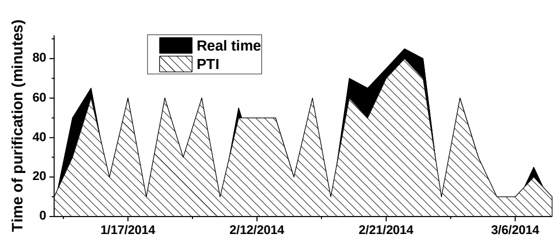


Figure 7. Indoor PM2.5 in a long period

EVALUATIONS

Datasets

In the evaluation, we use a real dataset of 150 workdays from 12/23/2013 to 5/9/2014 generated in Beijing campus:

1) *Indoor air quality records*: We collect indoor PM2.5 concentration every 10 minutes from our monitoring system.

2) *Outdoor air quality record*: We collect hourly outdoor PM2.5 concentration reported by the nearest air quality monitor station built by governments.

3) *Meteorological data*: We collect hourly fine-grained meteorological data from official websites, consisting of temperature, humidity, barometer pressure, and wind speed.

Baselines and Ground Truth

We compare our approach with four baselines:

1) Default: We choose the longest purification time (2 hours) in history as a default period.

2) *Average*: We set the average time of the historical PTs to reduce indoor PM2.5 concentration to a safe range.

3) *Regression*: A linear regression is employed to estimate the purification time, considering outdoor/indoor PM2.5 concentrations and meteorological data.

4) *ANN*: We only consider indoor and outdoor PM2.5 concentrations as the input of the PTI model.

Ground Truth: The data of the first two hours after turning on the HVAC, i.e., 5am–7am, is used in our experiments. Each two-hour time slot contains 12 records (one per 10 minutes). Regarding each of the record as a hypothetical beginning time, we obtain 12 instances of real purification time in a two-hour slot. We select data in the workdays (the HVAC is usually shot down in weekends), containing 733 instances from the 150-day dataset (we lost the data of some hours due to the failure of data collection). A 10-fold cross validation was employed to test the PTI model.

Results

We compare PTI with the four aforementioned baselines in Figure 8. Note that the inference is considered correct if the inferred PT equals to or is longer than the ground truth. As shown in Figure 8 A), the default period (2 hour) achieves a perfect accuracy, however, resulting in an over long PT, which wastes unnecessary energy. With a minor decrease in accuracy, our PTI model infers a much shorter purification time than the Default, therefore saving energy significantly, as depicted in Figure 8 B). PTI also has a shorter PT than Regression and Average and a similar PT as ANN. But, PTI outperforms these three baselines in term of accuracy.

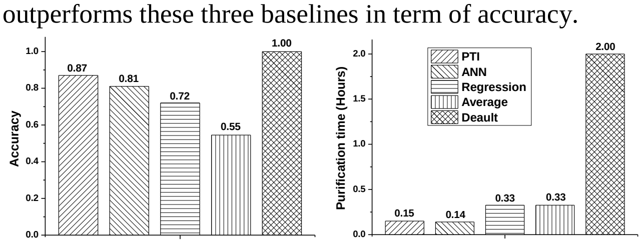


Figure 8. Overall results of different methods

RELATED WORK

There is a series of research on detecting indoor air quality. [5] described a personalized mobile sensing system MAQS to monitor CO2 concentration in a single room. [11] proposed a hybrid sensor network which contains both stationary sensors and mobile sensors to minimize the prediction error of the indoor CO2 concentration. However, actionable suggestion is not given in these research works.

Future indoor air quality can be predicted based on the sensed data. [4] presented an approach to predict indoor air pollution generated by cookstove emissions using a Monte Carlo model. Other mathematical models [8] are proposed for predicting indoor air quality based on smoking activity. Different from these methods, our approach considers more information, such as meteorological features and outdoor air quality, for a better prediction of purification time.

Various HVAC control strategies of smart buildings have been investigated in [1][2][6][7]. [1] presented a control architecture using sensing to guide operation of HVAC. [2] proposed a methodology with four phases to understand the energy performance and develop HVAC control scenarios to minimize energy usage. [6] used simulation models to verify against the effect of their strategies. [7] developed a multi-objective genetic algorithm which is validated using mathematic and simplified HVAC system problems. Different from these projects that focus on operating HVAC systems in an energy-effective way, we emphasize more on the integration of multiple data sources for a better prediction of PT. The latter is a typical approach in urban computing [13], which aims to solve the challenges in cities by using big data. In addition, we do not intervene the operating process after an HVAC system starts working. We just calculate the most energy-effective time to turn on an HVAC system ahead of its original schedule.

CONCLUSION

In this paper, we introduce an indoor air quality monitoring system deployed in four Microsoft campuses in China. The information of indoor air quality provided by the system can inform people's decision making in office areas. The gap between indoor and outdoor air quality can be used to measure the effectiveness of an HVAC in filtering air pollutants. The system also integrates outdoor air quality information with indoor measurements to adaptively control HVAC settings with a view on optimizing runtimes w.r.t. the energy efficiency and air quality conservation. Using a neural network-based approach, the time period that an HVAC needs to reduce the concentration of indoor PM2.5 into a healthy range is predicted based on six factors, such as the concentration of outdoor PM2.5 and humidity. Extensive experiments using 150-day data demonstrate the advantage of our approach beyond baseline methods, e.g., linear regression and average time. In addition, the meteorological features improves the accuracy of the prediction. With a minor decrease in accuracy, PTI infers a shorter purification time, thus saving energy significantly. We have released the data and execution file at [15].

REFERENCES

1. Agarwal, Y., Balaji, B., Dutta, S., Gupta, R. K., & Weng, T. (2011). Duty-cycling buildings aggressively: The next frontier in HVAC control. In *Information Processing in Sensor Networks (IPSN), 2011 10th International Conference on (pp. 246-257).* IEEE.
2. Canbay, C. S., Hepbasli, A., & Gokcen, G. (2004). Evaluating performance indices of a shopping centre and implementing HVAC control principles to minimize energy usage. *Energy and Buildings*, 36(6), 587-598.
3. Hasenfratz, D., Saukh, O., Sturzenegger, S., and Thiele, L.. Participatory Air Pollution Monitoring Using Smartphones. In *the 2nd International Workshop on Mobile Sensing*.
4. Johnson, M., Lam, N., Brant, S., Gray, C., & Pennise, D. (2011). Modeling indoor air pollution from cookstove emissions in developing countries using a Monte Carlo single-box model. *Atmospheric Environment*, 45(19), 3237-3243.
5. Jiang, Y., Li, K., Tian, L., Piedrahita, R., Yun, X., Mansata, Lv, Q.,… Shang, L. (2011, September). MAQS: a personalized mobile sensing system for indoor air quality monitoring. In *Proceedings of the 13th international conference on Ubiquitous computing* (pp. 271-280). ACM.
6. Mathews, E. H., Botha, C. P., Arndt, D. C., & Malan, A. (2001). HVAC control strategies to enhance comfort and minimise energy usage. *Energy and Buildings*, 33(8), 853-863.
7. Nassif, N., Kajl, S., & Sabourin, R. (2005). Optimization of HVAC control system strategy using two-objective genetic algorithm. *HVAC&R Research*, 11(3), 459-486.
8. Ott, W. R. (1999). Mathematical models for predicting indoor air quality from smoking activity. *Environmental Health Perspectives*, 107(Suppl 2), 375.
9. Pérez-Lombard, L., Ortiz, J., & Pout, C. (2008). A review on buildings energy consumption information. *Energy and buildings*, 40(3), 394-398.
10. Vardoulakis, S., Fisher, B. E. A., Pericleous, K., Gonzalez-Flesca. N., Modelling air quality in street canyons: a review. *Atmospheric Environment*, 37 (2003), 155-182.
11. Xiang, Y., Piedrahita, R., Dick, R. P., Hannigan, M., Lv, Q., & Shang, L. (2013, May). A Hybrid Sensor System for Indoor Air Quality Monitoring. In *Distributed Computing in Sensor Systems (DCOSS), 2013 IEEE International Conference on* (pp. 96-104). IEEE.
12. Zheng Y., Liu F, Hsieh H. U-Air: When Urban Air Quality Inference Meets Big Data. In *Proc. of KDD* 2013.
13. Zheng Y**,** Licia Capra, Ouri Wolfson, Hai Yang. Urban Computing: concepts, methodologies, and applications. *ACM Transaction on Intelligent Systems and Technology* (2014), 5(3).
14. GB3095. Ambient air quality standards [S]. Diss. 2012.
15. Download the indoor air quality data and execution files: http://research.microsoft.com/apps/pubs/?id=217237
16. Urban Air (Windows phone app download) http://www.windowsphone.com/s?appid=f36d5a33-2ccc-45f5-afd2-0c1afc5fc6dc