This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2881240, IEEE Internet of Things Journal

A Crowdsource-Based Sensing System for Monitoring Fine-grained Air Quality in Urban Environments

Jingchang Huang, Ning Duan, Peng Ji, Chunyang Ma, Feng Hu, Yuanyuan Ding, Yipeng Yu, Qianwei Zhou, Wei Sun

***Abstract*—Nowadays more and more urban residents are aware of the importance of the air quality to their health, especially who are living in the large cities that are seriously threatened by air pollution. Meanwhile, being limited by the spare sense nodes, the air quality information is very coarse in resolution, which brings urgent demands for high-resolution air quality data acquisition. In this paper, we refer the real-time and fine-gained air quality data in city-scale by employing the crowdsource automobiles as well as their built-in sensors, which significantly improves the sensing system's feasibility and practicability. The main idea of this work is motivated by that the air component concentration within a vehicle is very similar to that of its nearby environment when the vehicle's windows are open, given the fact that the air will exchange between the inside and outside of the vehicle though the opening window. Therefore, this paper firstly develops an intelligent algorithm to detect vehicular air exchange state, then extracts the concentration of pollutant in the condition that the concentration trend is convergent after opening the windo- ws, finally, the sensed convergent value is denoted as the equivalent air quality level of the surrounding environ- ment. Based on our IoT cloud platform, real-time air quality data streams from all over the city are collected and analyzed in our data center, and then a fine-gained city level air quality map can be exhibited elaborately. In order to demonstrate the effective-ness of the proposed method, experiments crowdsourcing 500 floating vehicles are conducted in Beijing city for three months to ubiquitously sample the air quality data. Evaluations of the algorithm's performance in comparison with the ground truth indicate the proposed system is practical for collecting air quality data in urban environments.**

***Index Terms*—CrowdSource Sensing, Vehicle Networking, Air Quality, Air Exchange State, Urban Environments**

I. INTRODUCTION

Nowadays, many industrial and daily activities generate huge air pollutions, especially in metropolitan cities of the developing countries, such as China and India. In these cities, air quality of the urban environment (e.g. the concentration of PM2.5) has seriously degraded people's work and life [1]. For example, Beijing, capital city of China, has to release red alarm of air pollution and stop nearly all of the public outdoor activities several times per year. Moreover, World Health Organization (WHO) reveals the fact that only in 2012, 875 thousand people died of air pollution, which contributes to 12.5% of all the death, indicating that the air pollution is now the largest single environmental health risk [2]. Meanwhile, air quality problem has attracted considerable amount of attentions from all over the world, and how to monitor air quality, discover outburst pollution event, estimate dispersion pattern and finally propose appropriate pollutions control methods are becoming hot research topics [3-6].

In order to study the air quality of an urban environment, the first step is to densely sample air components in each corner all over the city. The existing method of monitoring air quality, in most countries, is through the usage of static air pollution monitoring stations. These reference stations can provide highly accurate measurements from a limited number of specially selected sites, which should be representative of different types of locations [7]. Since these types reference stations are expensive, large and power hungry, so its number is very limited, e.g. only 23 stations are built in the whole Beijing city, whose area size is up to 50*50 km^2. In other words, in Beijing, each station should cover an average area of 113km^2, which makes the collected data sparse and must be interpolated with dedicated values [8].

Most of the work on urban environment air quality sensing fall into two categories: fixed station sampling and mobile network collecting [9]. Because the node number is limited, fixed station-based methods need to employ the interpolated results of the physical-chemistry process of raw data [8]. While mobile network collection schemes utilize sensors attached on mobile devices to collect data with geographic diversity. For example, OpenSense, a project run by EPFL and ETH in Zurich, aims to investigate mobile sensing techno-logies to monitor air pollution, however, only 10 buses equipped with air quality sensors are employed by this project to collect the air data around the city [10]. Therefore, no matter the solution based on fixed station or the mobile network based solutions, being limited by the cost of adding hardware, the number of sensor node is insufficient which leads a coarse resolution in spatial and temporal space. Moreover, there are lots of stochastic air pollution events around city, such as construction site, burst fire, traffic accidents and chemical leakage, which are valuable resources, but are not covered by the current sampling system.

High-resolution air quality data of urban environment is critical for city management, but how to collect air quality data in any place and at any time with a relatively low cost is still an unresolved problem [11]. Zheng [8] points out participatory sensing [12-14] may be potential solution to solve this problem in the future, if every person can carry a gas-sensor-equipped smart phone to probe the air quality arou-nd them [15]. Although this approach is feasible for some gasses like CO_2, it is not practical for other air pollutants like PM2.5 and TVOC so far, as the devices for sensing such kinds of air pollutants are heavy [16]. In this work, we fill this critical gap by presenting CrowdsourceSense, a crowdsource-

Jingchang Huang et. al, were with IBM-Research China, Shanghai, China. They were now research staff members with the interesting focus in the field of cognitive internet of things (e-mail: jchhuang@mail.ustc.edu.cn).

Copyright (c) 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

2327-4662 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2881240, IEEE Internet of Things Journal

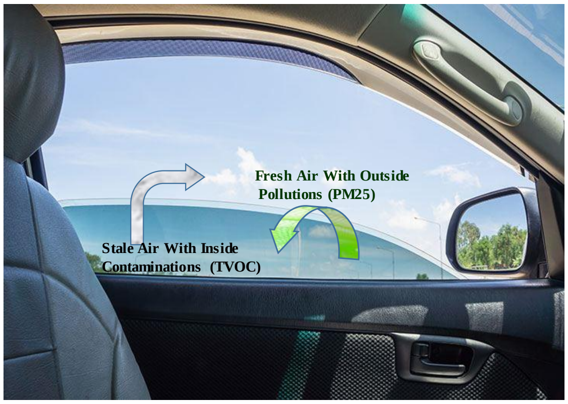

Fig. 1. Illustration of the air exchange between inside and outside vehicle when windows are open.

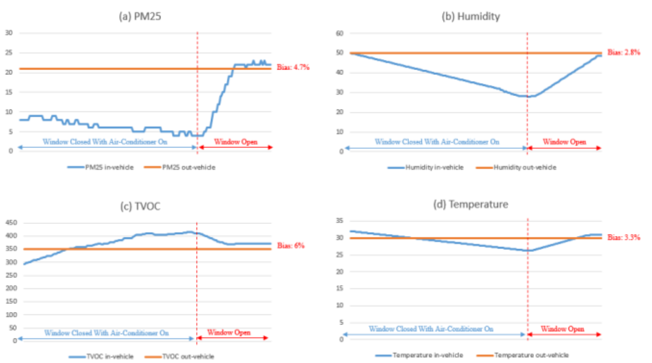

Fig. 2. Comparison of the air quality convergent value within vehicle and the quality of nearby outside environment.

based urban air quality sensing system, which leverages the quality sensors to collect data in fine-grained spatial and temporal resolution. The main idea of Crowdsource-Sense is motivated by the fact that the air component concentration in local urban environment is very close to that in vehicle when windows are open, given the air will exchange between inside and outside of a vehicle, as illustrated in Figure 1. Specifically, Figure 2 shows the trend of two sequential states, the state of window closed with air conditioner on and the state of window open, as we can see the concentration convergent value of window open segment is very close to the outside urban environment data with a maximal bias less than 6%. Therefore, in order to quantify the ambient humidity and temperature as well as the concentration of PM2.5 and TVOC, which are two of the most concerned air pollutants in urban environment, CrowdsourceSense firstly develops an intelligent algorithm to detect vehicular air exchange state, then extracts the concentration of pollutions, in the condition that the concentration trend is stable after opening the vehicular window, finally, the sensed stable concentration is denoted as the equivalent level of outer ambient environment. Because of the serious air pollution in big cities, more and more drivers are aware of the necessity to add air quality monitoring sensors and purifiers in their vehicle, resulting in the availability of air quality sensors data from lots of vehicles. In order to maintain healthy air quality in vehicle [17-19], more and more drivers are aware of the necessary to frequently open windows for air exchange, therefore, it is possible and feasible to employ the existing in-vehicle air quality sensors to monitor urban environment, which empowers the proposed system with the capability of sensing urban air quality ubiquitously and pervasively. The main contribution of this work is propose and implement the idea of using the opening window state (time slot) and existing built-in air sensors to capture outer air quality and finally achieving the aim of sampling data at any time and in any place.

The rest of the paper is organized as follows. Section 2 describes our system architecture, including the App, the sensors in the front-end and the IT system in the back-end. Section 3 illustrates the algorithm of how to sample data. Section 4 explains real data experimental results, and Section 5 gives a conclusion and points out some future research directions.

II. CROWDSOURCESENSE OVERVIEW

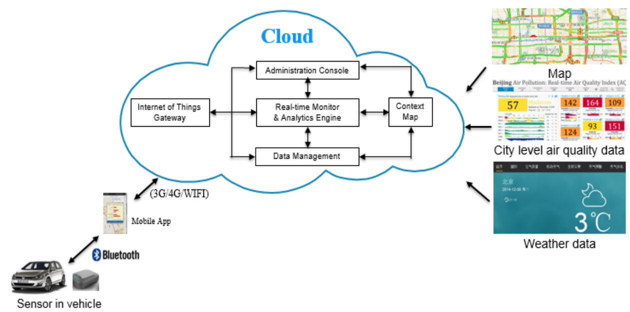

Fig. 3. System architecture of CrowdsourceSense

Figure 3 provides an overview of the system architecture of CrowdsourceSense. As illustrated, CrowdsourceSense consists of three principle components: a mobile App, an air quality sensing device and a cloud-based Internet of Things (IoT) platform [20]. The mobile App is the user interface for visualizing data and displaying interaction buttons. The air quality sensing device, namely the existing air sensors in vehicle, is used for quantifying the concentration of air components. The IoT platform is the management system of data stream, including IoT gateway, data management, context map, administration console, and real-time monitor and analytics engine. In detail, IoT gateway provides the capabilities of collecting streaming data from the mobile App, supporting batch data transmission and enabling the notification push from the platform to the users' mobile App. Data management module is responsible of managing all data transmitted to cloud. Context map is in charge of visualizing the aggregated data sources on the map, including traffic, weather, city level air quality data, road structure attributes, etc. Real-time monitor and analytics engine provides the streaming data processing, statistic and machine learning models to discover the user's air pollution footprints, profile and benchmark, as well as the air quality correlation patterns with contextual factors, such as traffic or weather [21]-[22].

In summary, this platform collects data from the sensor, mobile app and other data sources, such as weather, traffic and map for all kind of analytics demand, and is regarded as the tool for idea verification.

III. ALGORITHM DESIGN

2327-4662 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2881240, IEEE Internet of Things Journal

In order to sample air quality data anytime and anywhere, the key point is to detect the vehicular air exchange state (e.g. window open or closed), and then regard the pollutants concentration in vehicle as the local urban environment's value at the condition that there are air exchange between inside and outside vehicle, and the value of pollution concentration in vehicle is stable. Here, we define a vehicle is in open state if one or more of its windows are open and no matter how many inches each window is open. Although few new type of vehicles have the window state sensor and open the access to read it though on board diagnostic (OBD), however, most of the existing vehicles have no such configuration, moreover, this new function is impossible to be mature in near few years because the longer research and develop period in automobile industry, therefore, it is necessary to develop an intelligent detection algorithm to obtain air exchange state information. An overview of the proposed CrowdsourceSense is shown in Figure 4, where "out-loop" means air circulation through pulling outside air into a car's cabin (namely air outside mode), while "in-loop" represents air replacement by recirculating air already in the closed cabin environment (namely air inside mode). In our scheme, the first step is to separate one segment from another, then recognize each segment's air exchange state, finally, extract outside air quality from the estimated status of window open or window closed with air conditioner is in out-circulation mode, because only these two states' air quality can be close to local outside environment's.

A. *Air Quality Signature vs. Air Exchange State*

Most of the air pollution components in vehicle fall into two categories: generated in-vehicle and generated outside-vehicle. Pollution components generated in vehicle, for example, total volatile organic compound (TVOC), are produced by the inner ornaments, leather and perfume *et al*. Alternatively, air pollutions created by contaminations outside vehicle can circulate into vehicle through air exchange. Since there are pollutions exchange between inside and outside vehicle environment, as shown in Figure 1, the states, including window open, window closed, window closed with air conditioner in-loop and window closed with air conditioner out-loop, with the capability of inducing or hindering air circulation in vehicle, can significantly affect the concentration value of all contamination.

There are more than ten kinds of pollution components in air environments, among which PM2.5 and TVOC are the two most concerned air pollutants. Moreover, PM2.5 and humidity, respectively, have strong correlation relationship with states change of window and air conditioner, while they are also insensitive to other interferences factors, for example speed. Therefore, this work employs PM2.5 and humidity to segment the state change of window and air conditioner, then estimates the concentration value of PM2.5 and TVOC of local environment based on the states with air exchange between inside and outside vehicle.

Figure 5 illustrates the change signatures of PM2.5 and humidity in four different states: window open, window closed, window closed with air conditioner in in-loop mode, and window closed with air conditioner in out-loop mode. We conclude the following observations from each subfigure:

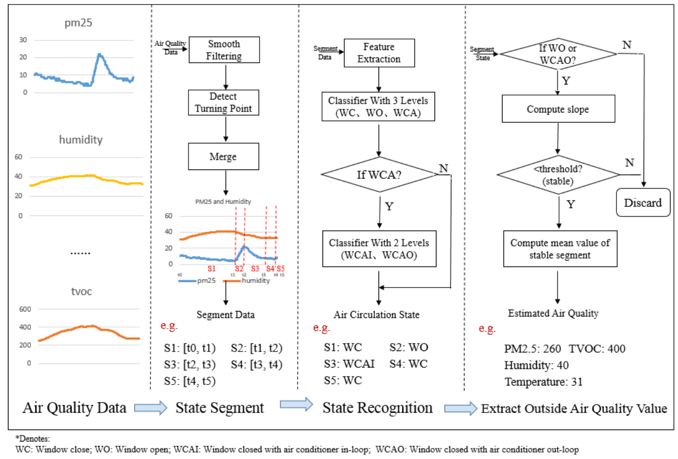

Fig. 4. Overview of the proposed algorithm architecture.

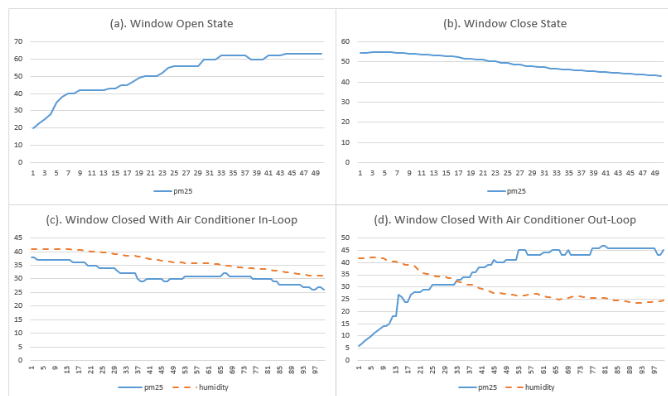

Fig. 5. Air quality signature vs. vehicular window and air conditioner states.

a. Window Open State (WO): PM2.5 concentration increases when a window is open, because the PM2.5 pollution source is from outside vehicle and it will circulate into vehicle though the open window.
b. Window Close State (WC): PM2.5 will decrease in most cases, which is opposite to the previous WO state.
c. Window Closed With Air Conditioner In-Loop State (WCAI): humidity decreases significantly since the air conditioner with the function of removing moisture from within the vehicle. At the same time, PM2.5 will decrease, similar with WC status.
d. Window Closed With Air Conditioner Out-Loop State (WCAO): Similar with the WCAI, the humidity of this state will decrease. The PM2.5 value is uncertain and depends on vehicle usage, air conditioner configuration and traffic congestion.

B. *Window State Segmentation*

After observing the change trend of the sample data in Figure 2, we know that the turning points of PM2.5 sequence and humidity sequence indicate the states change of the window and air conditioner, so this phenomenon can be further utilized to separate one state from another. Specifically, on one hand, both the state changes of window and air conditioner will stimulate strong response of PM2.5, while only air conditioner can stimulate humidity varying regularly. On the other hand, the response of humidity to air conditioner is more

2327-4662 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2881240, IEEE Internet of Things Journal

sensitive and accurate than PM2.5, so the analysis of state change based on humidity is more accurate than PM2.5. Therefore, it is reasonable to detect state change though combining the values of PM2.5 and humidity. The reason why TVOC and temperature are not used for segmentation is that they are easy be disturbed by interferences, for example, TVOC will be affected by speed variation while temperature will be contaminated by illumination.

In order to control cost, mobile devices usually adopt consumer-grade air quality sensors, whose resolution is coarse besides performance is unstable, thus results with noisy data. Undoubtedly, the interferences will strongly degrade the accuracy of the segmentation process. Exponential smoothing (ES) is commonly applied to smoothen data, acting as low-pass filters to remove high frequency noise [20]. In ES filtering, the raw data sequence is often represented by x_t beginning at time t = 0, and the output of the exponential smoothing algorithm is commonly written as s_t, which may be regarded as a best estimate of what the next value of x will be. When the sequence of observations begins at time t = 0, the simplest form of exponential smoothing is given by the Equation (1), where α is the smoothing factor, and 0 < α < 1.

$$\begin{cases} s_0 = x_0 \\ s_t = \alpha x_t + (1-\alpha)s_{t-1}, \qquad t>0 \end{cases} \tag{1}$$

Moreover, it is easy to ensemble multiple ES filters to deal with signals having different signal-noise ratio. Therefore, in this paper, ES filter is first adopted to preprocess air quality data to remove the burst interferences introduced by the sensor instability. In our experiment, ES filter with 3-order is configured for PM2.5 preprocessing while only 2-order is for humidity since the data quality of humidity is a little better than PM2.5.

Based on the signatures observed in Figure 5, we define the separation criteria as follows:

a) Change trend of a curve (PM2.5 or Humidity) reverses;
b) Cumulative change rate exceeds a threshold obtained from training dataset;

Once one of the above conditions occurs, a segmentation decision is made, the detailed workflow of state segment process is presented in **Algorithm 1**. If the interval of two adjacent segments is too short, we merge them into one, and use the middle time as the time of segmentation. A detailed visualization of the ES filtering and segmentation is shown in Figure 6.

Algorithm 1: State Segmentation

Input: Real-time air quality data, including PM2.5 and Humidity
Output: Segment flag

```
i ⟵ 0
Do
    segment( i ) ⟵ false
    humi_rate( i ) ⟵ humi( i ) - humi(i-1);
    PM25_rate( i ) ⟵ PM25(i) - PM25(i-1);
    If ((humi_rate( i ) * humi_rate(i-1))<0) || ((PM25_rate( i )* PM25_rate(i-1)) <0)
        segment ( i ) ⟵ true;
        Break;
    End
    If ( ∑_i^{i-N} (humi_rate(i)) > humi_threshold ) || ( ∑_i^{i-N} (pm25_rate(i)) > pm25_threshold )
        segment ( i ) ⟵ true;
        Break;
    End
    return segment( i );
    i ++;
Until trajectory end
```

C. Feature Extraction

There are few existing datasets at project initial stage, so the deep learning networks which have the characteristic of blind feature are not feasible for this case. In order to recognize the state of each segment, the corresponding feature needs to be extracted. Since the previous state will influence the feature of the current state, (e.g., the PM2.5 will decrease when current state is WC, however, the PM2.5 will increase if the current state is WC and its precious state is WCAO, which is opposite to most situations), therefore, the extracted feature should have the capability of descripting the signatures of both current and previous states. Here, the information of two sequential states are employed for recognition, specifically, the endpoints of previous state and all data points of the current state are used for feature extraction. The reason why only the endpoints of previous segment are employed for feature extraction rather than all data points, is that the storage of all data points consumes too much memory. The illustration of data employed for feature extraction is shown in Figure 7. The specific feature types employed in our experiment and their corresponding physical meanings are listed in Table I. Finally, the feature vector, consisting of both the previous segment and current segment information are input into the classifier.

$$\mathbf{Feature} = \{\mathbf{preFeature}, \mathbf{curFeature}\} \tag{2}$$

$$\mathbf{preFeature} = \{alpha\_PM2.5, alpha\_TVOC, alpha\_Humi, alpha\_Temp\} \tag{3}$$

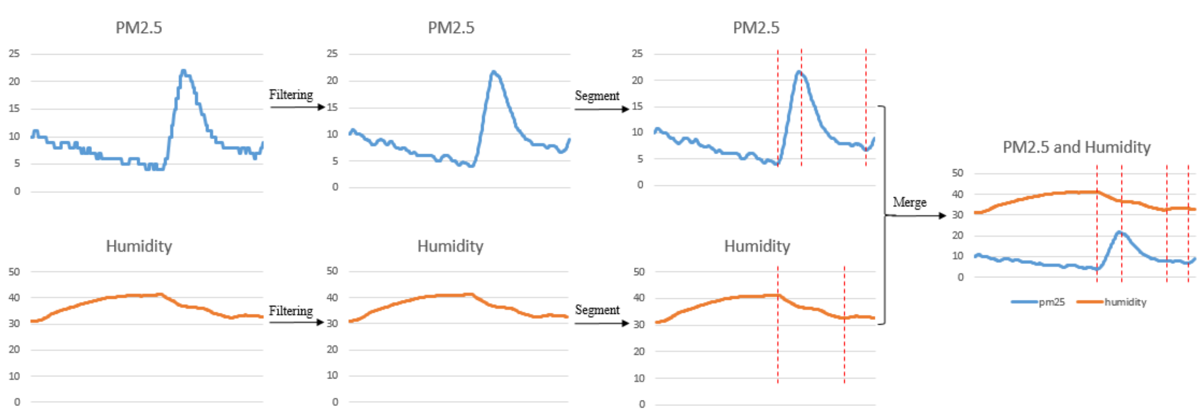

Fig. 6. The illustration of segmentation.

2327-4662 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2881240, IEEE Internet of Things Journal

$$\textbf{curFeature} = \{PM2.5\_rate1, PM2.5\_rate2, Humi\_rate1, \cdots, CorrTVOCHumi\} \quad (4)$$

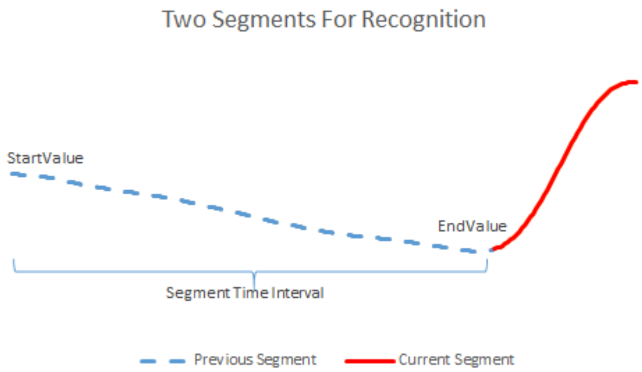

Fig. 7. Illustration of data usage for feature extraction.

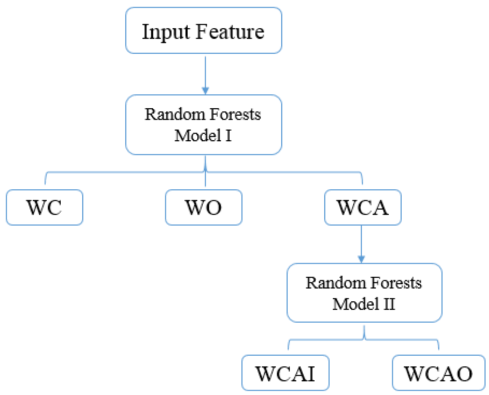

Fig. 8. Classification strategy

D. State Classification

The air quality in vehicle is influenced by variant environmental factors, such as weather, traffic, windows and air conditioner switch status, and vehicular air conditioner filter maintenance statue (new or worn), which induces complexity and nonlinear on feature. Therefore, rule-based classifiers or other light-weight classifiers cannot deal with it. In this paper, the random forest algorithm is employed to classify the air circulation states. Random forests (also known as random decision forests) is an ensemble learning method for classifycation, regression and other tasks, that operate by constructing a multitude of decision trees at training stage and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees [21]. Random decision forests correct decision trees' habit of over-fitting to their training set.

In order to improve the classification accuracy, a hierarchical classification strategy with two random forests models is employed for air circulation state recognition which overcomes the confusion of feature space between WC and WCAI, WO and WCAO. The diagram of the classification model is shown in Figure 8.

IV. EXPERIMENT

Experimental studies were performed on all urban districts around Beijing city, China, to demonstrate the feasibility and practicality of our CrowdsourceSense system. Specifically, there are 500 crowdsourcing vehicles employed for data collection and each vehicle is equipped with an air quality sensing device (as shown in Figure 9), a mobile application, and a supporting Cloud-based IoT platform. The whole experiment lasts three months, and runs over 1.3 million kilometers mileages trajectory, which corresponds to around 2.41 driving hours per day per driver. While the data being collected, each vehicle is operated by at least two persons, one is responsible for driving and the other takes charge of recording air circulation state and its corresponding time stamp. In order to evaluate the performance of the air circulation state recognition algorithm, half of the datasets are used for training and the other half for testing. The testing results are detailed as table II.

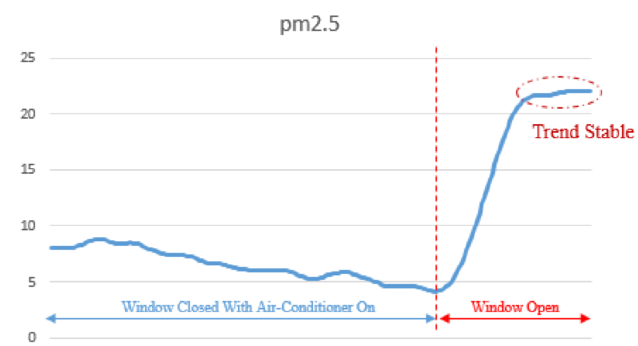

Fig. 10. Illustration of regarding air quality in-vehicle as urban environment.

In table II, each row denotes the actual air circulation state and each column represents the results recognized by CrowdsourceSense. According to table II, we know that the states with air conditioner on (WCAI and WCAO) outperform the

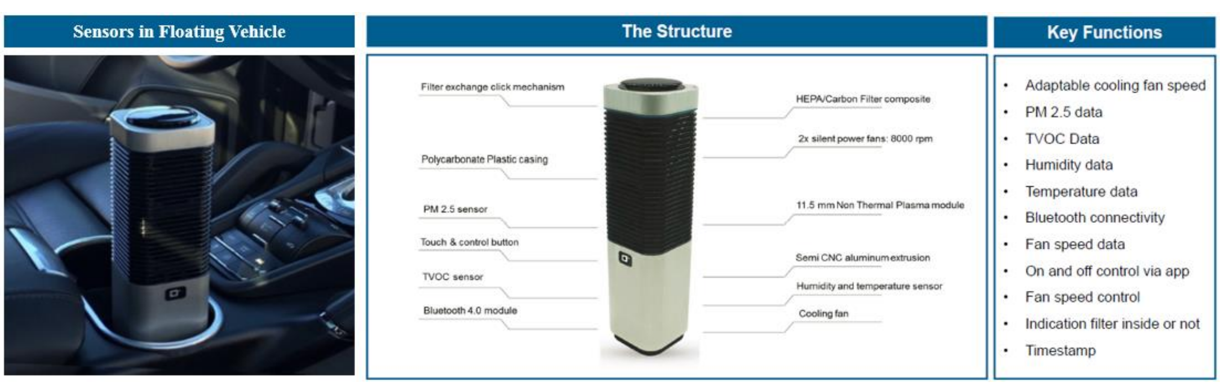

Fig. 9. Illustration of sensors in experimental vehicle

2327-4662 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2881240, IEEE Internet of Things Journal

Table I. Feature list and the corresponding physical meanings

Feature List			
Segment	Feature	Formula	Description
Previous Segment	alpha_PM2.5	alpha_PM25= $\frac{PM25EndValue-PM25StartValue}{SegmentTimeInterval}$	Slope of PM2.5 of previous segment.
	alpha_TVOC	alpha_TVOC= $\frac{TVOCEndValue-TVOCStartValue}{SegmentTimeInterval}$	Slope of TVOC of previous segment.
	alpha_Humi	alpha_Humi= $\frac{HumiEndValue-HumiStartValue}{SegmentTimeInterval}$	Slope of humidity of previous segment.
	alpha_Temp	alpha_Temp= $\frac{TempEndValue-TempStartValue}{SegmentTimeInterval}$	Slope of temperature of previous segment.
Current Segment	pm2.5_rate1	$pm25\_rate1 = \sum_{i=2}^{i=Length}\left(\frac{pm25(i)-pm25(i-1)}{PointTimeInterval}\right)$	Accumulation of variation rate of PM2.5.
	pm2.5_rate2	$pm25\_rate2 = \sum_{i=2}^{i=Length}\left(\frac{pm25(i)-pm25(i-1)}{pm25(i-1)}\right)$	Accumulation of relative variation rate of PM2.5.
	humi_rate1	$humi\_rate1 = \sum_{i=2}^{i=Length}\left(\frac{humi(i)-humi(i-1)}{PointTimeInterval}\right)$	Accumulation of variation rate of humidity.
	tvoc_rate1	$tvoc\_rate1 = \sum_{i=2}^{i=Length}\left(\frac{tvoc(i)-tvoc(i-1)}{PointTimeInterval}\right)$	Accumulation of variation rate of TVOC.
	tvoc_rate2	$tvoc\_rate2 = \sum_{i=2}^{i=Length}\left(\frac{tvoc(i)-tvoc(i-1)}{tvoc(i-1)}\right)$	Accumulation of relative variation rate of TVOC.
	speed_rate1	$speed\_rate1 = \sum_{i=2}^{i=Length}\left(\frac{speed(i)-speed(i-1)}{PointTimeInterval}\right)$	Accumulation of variation rate of speed.
	Mean_pm2.5	$Mean\_pm25 = \sum_{i=1}^{i=Length}\left(\frac{pm25(i)}{Length}\right)$	Mean value of PM2.5.
	Corr_PM2.5Speed	$Corr\_PM25Speed = Correlation(\mathbf{PM25}, \mathbf{Speed})$	Correlation coefficient between PM2.5 and Speed.
	Corr_TVOCSpeed	$Corr\_TVOCSpeed = Correlation(\mathbf{TVOC}, \mathbf{Speed})$	Correlation coefficient between TVOC and Speed.
	Corr_TVOCHumi	$Corr\_TVOCHumi = Correlation(\mathbf{TVOC}, \mathbf{Humi})$	Correlation coefficient between TVOC and humidity.

states with air conditioner off, benefiting from the aids of humidity feature. Additionally, respectively, the results of the the states with air conditioner off, benefiting from the aids of humidity feature. Additionally, respectively, the results of the states with window close (WC and WCAI) are better than those of window open (WO and WCAO) respectively, since the number of inches a window opens can fluctuate the change rate of air components, so the features of window close state are more stable.

Table II Confusion matrix of the air circulation state recognition of CrowdsourceSense

	WC	WO	WCAI	WCAO	Recall (%)
WC	22642	941	613	572	91.4
WO	792	23175	680	998	90.4
WCAI	321	200	32909	1571	94.0
WCAO	371	272	1011	23157	93.3
Precision (%)	93.8	94.3	93.5	88.1	

After recognizing air circulation states, we employ the following strategies to collect urban environment air quality data:

a) Regard the in-vehicle air quality as urban environment's measured value in the condition that vehicular window is open as well as the concentration trend is stable, as Figure 10 shows.
b) For the similar signatures between WC and WCAO, replace WC with WCAO to execute the strategy a) for the regions where the number of WC segment is few.

Figure 11 shows the graphic interface of datasets collection, which not only illustrates the location distribution of the floating vehicles, but also reports each vehicle's real-time air circulation state and the obtained air quality concentration. In this figure, the color of crowdsourcing vehicle icon represents the pollution level. A legend showing the real-time air data and air exchange state will appear when clicking on the vehicular icon. Since the distribution of the crowdsourcing vehicles cannot cover all areas of a city, the interpolation operations are necessary for getting all positions' air quality, this paper employ the method proposed by Zheng [8] to deal with it. To verify the validity of our sample method, we regard the data reported by government public stations as the ground truth to evaluate the performances of CrowdsourceSense. Using this approach, the obtained air quality of the air station locating in central area of Beijing and the corresponding ground truth are compared. Here, we define the location whose position within the fifth ring of Beijing as downtown, and those outside of the fifth ring of Beijing as suburb. According to the experiment, we find that the the estimated results of downtown are higher than the ground truth while that of suburb is lower than that the ground truth, which can be explained by that traffic in downtown is more congestion and generates more particle matter than suburb. We also find that the estimated values of our proposed method are higher than that of public monitoring station. This is because the altitude of public air monitoring station is higher than those sensors attached in moving vehicle and the higher the altitude, the less

2327-4662 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2881240, IEEE Internet of Things Journal

dust are produced. Although there are differences between the estimated results and the ground truth, specifically, an average bias value of 3.64% is achieved and 4.2% bias is concluded from all datasets, nevertheless, the bias is within an acceptable threshold (5% required by funder), so the bias result convincingly demonstrates that the CrowdsourceSense has the ability to sample urban environment air quality.

V. DISCUSSION ON POTENTIAL APPLICATIONS AND WEAKNESS

The evaluation results verify that CrowdsourceSense is a promising data collection technology for urban computing since it is a fundamental function demanded by lots of applications. In this section, we discuss its potential applications and possible future work.

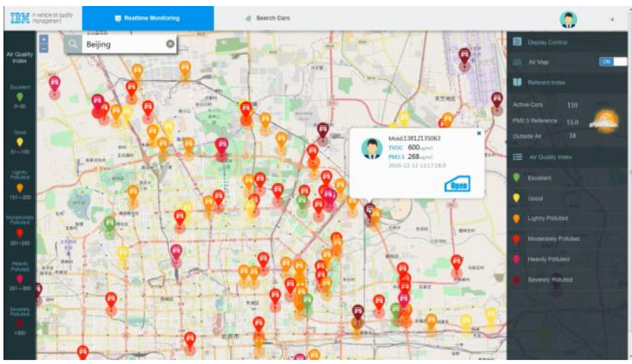

Fig. 11. CrowdsourceSense interface for data sampling

A. Build high-resolution dynamic air quality map

Nowadays, government only releases the air quality of a few sites, as mentioned before, in Beijing, only 23 monitoring stations are available to public, which means the residents in every 113km^2 area are sharing the same reference value. Therefore, the resolution of the current air quality information is coarse and needs improvement. By employing CrowdsourceSense to collect data, with the assistance of some advanced interpolation method, a real-time and high-resolution air quality map is available to the public. Based on this, not only local government can improve city management level, such as discovering pollution leakage event, optimizing energy conservation and emissions reduction, the city residents can also prevent respiratory disease through avoiding entering pollution area or adding protection measures in a timely manner.

B. Refine air dispersion physical model

The simulation, based on air dispersion physical model, is usually adopted to predict the air quality. However, the physical model is derived from pure theory, or even some simple superposition of several transdisciplinary formulas. Additionally, because of the lacking of real data, current physical model has not been clearly verified whether it is accurate or not. Fortunately, with the help of CrowdsourceSense, we can obtain both the air quality data of each grid in the monitoring area and the correlation relationship among environment, population and transport, which are the input and output for the verification of air dispersion model. Therefore, we think the CrowdsourceSense is a practical platform to verify all kinds of air dispersion models.

C. Estimate vehicle air circulation maintenance state

As known to all, air conditioner filter has the capability of cleaning air quality because it can remove the dust, bacteria and other particles from the incoming air. As time goes by, the air filter will accumulate more and more dirty matters, and finally lose its function of cleaning. Currently, drivers replace new air conditioner filter empirically, e.g. by replacing a new air conditioner filter every three months or every few thousands of mileage, which is not optimized since the decision is not related to the air quality. With the CrowdsourceSense, drivers can clearly know the real-time air values, the comparison results between himself/herself and other drivers as well as the ranking of his/her air quality in the whole city [22]. Based on the comprehensive information, drivers can make the filter replacement decision timely and accurately. Additionally, employing the big data obtained by CrowdsourceSense, auto manufacturer can further mine user behavior about air circulation, track air purify performance of current vehicle types and even aid the design process of the next generation automobile.

There are also a few limitations in CrowdsourceSense analytics algorithm. First, if there are no WO segments, this paper directly regard WCAO as WO to collect air data without considering the air quality deviation between WCAO and WO, which may bring some bias to the final results. Therefore, a revised model relating to position altitude and air filter maintenance status needs to be explored in near future. Second, current classifier is fully tested in only four cities in China (Beijing, Shijiazhuang, Zhengzhou and Chengdu), the former three are all seriously polluted places and the feature differences between window close and window open are significant, the last one Chengdu city is slightly polluted and with trivial difference between inside and outside vehicle. How about the distances among inter-classes of cities locating in developed countries that has excellent air quality? Can it be correctly recognized? More interesting results may be revealed if CrowdsourceSense are tested in more diverse environments.

VI. CONCLUSIONS

In this paper, from the perspective of collecting data ubiquitously, we propose a crowdsource-based urban air quality sensing system, which leverages the ubiquitous automobiles and their inside air quality sensors to collect data in fine-grained spatial and temporal resolution. Our solution has the capability of sampling air quality data in any place and at any time, and thus empower a data source for lots of urban computing applications, filling the critical gap between urban computing demands and the current situation of insufficient data source. Experiments involving 500 crowdsourcing vehicles for three months are conducted to evaluate the performance of CrowdsourceSense. Our results reveal that the proposed system can achieve 92.4% accuracy in window status recognition and less than 9% bias for estimating the air quality in Beijing. In near future, we plan to further improve the accuracy of the estimated air quality concentration by modifying the bias between the WC and WCAO, and explore more potential applications of the collected data for urban computing.

2327-4662 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2881240, IEEE Internet of Things Journal

VII. Acknowledge

We thank the anonymous reviewers of IEEE Internet of Things Journal for their valuable reviews and insightful comments. We are also grateful to DiDi Corporation [26] for providing help in scheduling drivers and experimental vehicles. This work was supported in part by Natural Science Science Foundation of China under Grant 61802347.

References

[1] J. Daisey, W. Angell, and M. Apte, Indoor air quality, ventilation and health symptoms in schools: an analysis of existing information. Indoor air, 2003. 13(1): p. 53-64.

[2] W.H. Organization, 7 million premature deaths annually linked to air pollution. World Health Organization, Geneva, Switzerland, 2014.

[3] A. Jutzeler, J.J. Li, and B. Faltings, A Region-Based Model for Estimating Urban Air Pollution. in AAAI. 2014.

[4] D. Hasenfratz, O. Saukh, C. Walser, and C. Hueglin, Deriving high-resolution urban air pollution maps using mobile sensor nodes. Pervasive and Mobile Computing, 2015. 16: p. 268-285.

[5] S. Devarakona, P. Sevusu, and H. Liu, Real-time air quality monitoring through mobile sensing in metropolitan areas. in Proceedings of the 2nd ACM SIGKDD international workshop on urban computing. 2013. ACM.

[6] Y. Cheng, X. Li, Z. Li, S. Jiang, and Y. Li, J JiaAirCloud: a cloud-based air-quality monitoring system for everyone. in Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems. 2014. ACM.

[7] A. Marjovi, A. Arfire, and A. Martinoli, High resolution air pollution maps in urban environments using mobile sensor networks. in Distributed Computing in Sensor Systems, 2015 International Con. on. 2015. IEEE.

[8] Y. Zheng, F. Liu, and HP Hsieh, U-air: When urban air quality inference meets big data. in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. 2013. ACM.

[9] Y. Yang, Z. Zheng, K. Bian, and L Song, Real-time profiling of fine-grained air quality index distribution using uav sensing. IEEE Internet of Things Journal, 2018. 5(1): p. 186-198.

[10] Project OpenSense II: Sensing the air we breathe. Available: http://opensense.epfl.ch.

[11] Y. Xiang, R. Piedrahita, and RP. Dick, A hybrid sensor system for indoor air quality monitoring. in Distributed Computing in Sensor Systems (DCOSS), 2013 IEEE International Conference on. 2013. IEEE.

[12] Y. Jiang, K. Li, L. Tian, R. Piedrahita, and X. Yun. MAQS: a personalized mobile sensing system for indoor air quality monitoring. in Proceedings of the 13th int. conf. on Ubiquitous computing. 2011. ACM.

[13] D. Hasenfratz, O. Saukh, S. Sturzenegger, and L. Thiele, Participatory air pollution monitoring using smartphones. Mobile Sensing, 2012: p. 1-5.

[14] L. Chen, Y. Ho, H. Hsieh and S. Huang, ADF: an Anomaly Detection Framework for Large-scale PM2. 5 Sensing Systems. IEEE Internet of Things Journal, 2018. 5(2): p. 559-570.

[15] A. Al-Ali, I. Zualkernan, and F. Aloul, A mobile GPRS-sensors array for air pollution monitoring. IEEE Sensors Journal, 2010. 10(10): p. 1666-1671.

[16] F. Tsow, E. Forzani, A. Rai, R. Wang, and R. Tsui, A wearable and wireless sensor system for real-time monitoring of toxic environmental volatile organic compounds. IEEE Sensors J., 2009. 9(12): p. 1734-1740.

[17] M.S. Alam and A. M., A critical review and assessment of Eco-Driving policy&tech.: Benefits & limitations. Trans. Policy, 2014. 35: p. 42-49.

[18] N. Hudda, E. Kostenidou, and C. Sioutas, Vehicle and driving characteristics that influence in-cabin particle number concentrations. Environmental science & technology, 2011. 45(20): p. 8691-8697.

[19] C. Ma, N. Duan, W. Sun, F. Hu, J. Huang, and P. Ji, Reducing air pollution exposure in a road trip. in 24rd World Congress on Intelligent Transportation Systems Montreal, Canada. 2017.

[20] W. Sun, N. Duan, and P. Ji, Intelligent in-vehicle air quality management: a smart mobility application dealing with air pollution in the traffic, in 23rd ITS World Congress. 2016: Melbourne, Australia.

[21] L. Cagliero, T. Cerquitelli, S. Chiusano, and P. Garza. Discovering air quality patterns in urban environments. in Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. 2016. ACM.

[22] Y. Zheng, X. Chen, Q. Jin, and Y. Chen, A cloud-based knowledge discovery system for monitoring fine-grained air quality. MSR-TR-2014–40, Tech. Rep., 2014.

[23] E.S. Gardner, Exponential smoothing: The state of the art. Journal of forecasting, 1985. 4(1): p. 1-28.

[24] A. Liaw, and M. Wiener, Classification and regression by randomForest. R news, 2002. 2(3): p. 18-22.

[25] S. Kim, and E. Paulos. InAir: sharing indoor air quality measurements and visualizations. in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010. ACM.

[26] DiDi Corporation. Available: http://www.didiglobal.com/#/.

Jingchang Huang received the Ph.D. degree from University of Chinese Academic and Sciences, China in 2015. He joined IBM Research-China in 2015 and currently he is a research staff member in Cognitive-IoT team. He mainly focuses on the cognitive data analytics, including acoustic, image and time sequences, which all comes from Internet-of-Thing related applications.

Ning Duan is a Senior Technical Staff Member of IBM Research-China leading connected vehicle related AI analysis and data mining research. In recent years, Mr. Duan has been driving innovation projects in the areas of smart urban mobility services, fleet management optimization, and intelligent transportation through partnership with clients in related industries.

Feng Hu obtained the PhD degree in Computer Science at The Graduate Center, CUNY in 2017. His research interests include computer vision, 2D/3D based indoor localization, argument reality/virtual reality, self-driving vehicles, machine learning, video surveillance, and parallel computing.

Peng Ji is a Research Staff Member in IBM Research-China, major in big data mining and computer vision AI analysis. In recent years, He had participated many projects on connected vehicle data analysis services for urban mobility, fleet management optimization and intelligent transportation.

Chunyang Ma received the Ph.D. degree in Computer Science from Zhejiang University, China in 2012. She joined IBM Research-China as a research scientist the same year. Her main research interests includes spatiotemporal data management, location based services and intelligent transportation.

Yipeng Yu received the Ph.D. degree in Computer Science from Zhejiang University in 2016. He joined IBM Research-china as a research scientist in the same year. His main research interests include pervasive computing, signal processing, natural language processing, and automatic machine learning.

Qianwei Zhou joined Zhejiang University of Technology, Hangzhou, China, in July 2014 where he is currently a research scientist in the College of Computer Science. His research interests include the crossing field of machine learning and computer aided design, the Internet-of-Thing related signal processing and pattern recognition.

Yuanyuan Ding received her Ph.D. degree in July 2017 from Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Science, China. She is now a research scientist in IBM. Her research interests are image processing, pattern recognition and moving object detection.

Wei Sun is currently leading the research and development activities of smart mobility initiatives of Volkswagen Group China. Dr. Sun is responsible to apply artificial intelligence, internet of things, big data and cloud technologies in building strategic and innovative new mobility services. He was a member of IBM Academy of Technology that is an IBM global top technical leadership community, and he was also a member of IBM Academy Leadership Team driving strategic global programs.

2327-4662 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.