Poster Abstract: Robust Calibration for Low-Cost Air Quality Sensors using Historical Data

Guodong Li
ligd19@mails.tsinghua.edu.cn
Tsinghua-Berkeley Shenzhen Institute
Tsinghua University

Xinyu Liu
liuxinyu13@mails.tsinghua.edu.cn
Dept. of Electronic Engineering
Tsinghua University

Zhiyuan Wu
wu-zy18@mails.tsinghua.edu.cn
Dept. of Electronic Engineering
Tsinghua University

Yue Wang
wangyue@tsinghua.edu.cn
Dept. of Electronic Engineering
Tsinghua University

Lin Zhang
linzhang@tsinghua.edu.cn
Tsinghua-Berkeley Shenzhen Institute
Tsinghua University

ABSTRACT

As pollution problems become increasingly prominent nowadays, urban air quality monitoring has attracted more and more attention. In recent years, sensing systems based on low-cost sensors are proposed to achieve fine-grained monitoring with larger amount of deployment as supplyment to conventional monitoring stations. Calibration is critical to guarantee the accuracy and consistency of these sensing systems to fight against sensor drift. While conventional field calibration approaches often rely on real-time data from a nearby standard station, they are not applicable to low-cost sensors which cannot receive the latest reference data from nearby stations after deployment. In reality, it is very difficult for sensors to get access to nearby standard stations deployed sparsely. To reduce the dependency on real-time and nearby reference data, we present a Robust Calibration approach based on Historical data (RCH) for the low-cost air pollution sensor calibration. Our method corrects the sensor drift by adapting sensitivity and offset based on estimating the probability distribution of pollutant's concentration. Experiments with real-world NO_2 data in Foshan, China show that our proposed method acheives close performance to conventional field calibration methods but addresses above challenges. Moreover, our method can use historical data collected from the sensors in more distant geographic locations than the compared method.

KEYWORDS

robust calibration, air pollution, low-cost sensor

1 INTRODUCTION

There are 4.2 million premature deaths are caused by air pollution every year according to WHO [5]. Air pollution and its monitoring have attracted a lot of attention. Conventional monitoring approaches deploy accurate stations, but these stations are deployed spatially sparsely due to high per-unit price. To support fine-grained air pollution monitoring, low-cost air quality sensors have been proposed and deployed to monitor air pollution[3]. Although these sensors can be deployed densely with acceptable total cost, they also suffer from measurement deviation due to unavoidable sensor drift. So sensor calibration is a necessity to ensure the data quality of low-cost sensors as well as the stability of the sensing systems.

Even with lab-calibration before deployment, sensor's behavior may change over time [4]. The calibration parameters of sensors including sensitivity and offset require adapting to its surrounding environment. It should be noted that each sensor's individual difference is stable for a long time, and experiments show that the correlation between the readings of the original sensor and the nearest official monitoring station is very strong [1], making deployed sensors' calibration possible without recalling.

In most calibration cases monitoring stations are used as the reference sources. But it is usually very difficult for sensors to get access to real-time reference data from nearby stations [2], especially when fixed-deployed sensors cannot be a 'neighbour' with any monitoring stations both temporally and spatially. As a result, how to calibrate sensors without real-time data from nearby stations as references becomes very challenging. To address it, we present a robust calibration approach based on historical data, an optimization based method to calibrate low-cost air pollution sensors. We make use of the similarity in distributions of references during a period of time to make it possible that sensors can be calibrated without dependency of stations' synchronized measurements after deployment.

2 ROBUST CALIBRATION BASED ON HISTORICAL DATA

Since the distribution of air pollution data changes slowly, we can find a reference source in the case where we have $p(X_r) \approx p(X_t)$ within the same time window of different days. $p(\cdot)$ denotes the probability distribution. X_r denotes the reference based on the historical data of the monitoring station and X_t denotes data from the target sensor to be calibrated. We can find the function F by solving the following optimization problem.

$$\mathrm{argmin}_F \; d_{KL}[\hat{p}(X_r),\ \hat{p}(F(X_t))] \tag{1}$$

Here, F denotes a linear function including the calibration parameters: sensitivity and offset. And $\hat{p}$ denotes a probability density estimator. And d_{KL} denotes the Kullback-Leibler divergence between two probability distributions, which can measure 'distance' between distributons. In this work, Gaussian Mixture Model (GMM) as the probability density estimator has been used to fit both the reference data and the target low-cost sensor data.

$$\hat{p}(x) = \sum_{i=1}^{k} \frac{A_i}{\sigma_i \sqrt{2\pi}} e^{-(x-\mu_i)^2 \big/ 2\sigma_i^2} \tag{2}$$

DOI 10.1109/IPSN48710.2020.00-12

where A_i denotes different weights for gaussian distributions. And σ_i is the standard deviation and μ_i is the mean. By solving the optimization problem (1), we can find the optimal calibration value of sensitivity and offset of the sensor to correct sensor drift. Our method is shown in Figure 1.

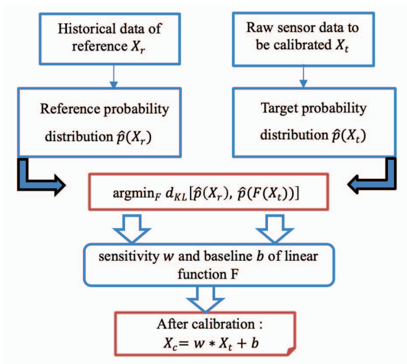

Figure 1: Robust calibration based on historical data

3 EVALUATION

In the experiment, we compare our RCH method with least-square (LS) calibration, a conventional field calibration method. The results show that without real-time and nearby reference, RCH can still have similar calibration effect compared with that using least-square calibration method.

3.1 Experiment Setup

We deployed the target sensor to be calibrated in Foshan, Guangdong province in China and we have observed it for 1 month in November. Gaussian Mixture Model (GMM) has been used to fit both the target sensor data (NO_2) to be calibrated in November and the monitoring station data used as reference in October, which is about 6 km away from the target low-cost sensor. We chose $k = 4$ in (2) because it is enough for having a good fitting of distributions. And we used RCH to get optimal calibration parameters sensitivity and offset of the sensor by minimizing the Kullback-Leibler divergence between the reference distribution and the target distribution. Then, we corrected the sensor drift with the sensitivity and offset obtained.

When deploying this target sensor, we also intended to do another experiment for comparing RCH with least-square calibration with real-time data. So we also deployed the target sensor next to another monitoring station with the distance about 20 m in order to do the least-square calibration with synchronized measurements of this nearby station.

3.2 Algorithm Performance

The least-square calibration is a conventional field calibration method which gets the real-time data from nearby monitoring station to calibrate the target sensor. The results of RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error) calculated by data in November of these two calibration methods are shown in Table 1. A typical example shows original values of NO_2 and values after calibration by RCH and least-square calibration in Figure 2. It is shown

Table 1: Calibration results of RCH and LS

calibration method	RMSE(ug/m^3)	MAE(ug/m^3)
before calibration	36.238	32.791
RCH	17.307	14.367
least-square calibration	16.640	13.668

that the performance of RCH is close to that using least-square calibration with real-time data. That is to say, even without any real-time reference data from nearby station, RCH can still have similar calibration effect compared with least-square calibration method.

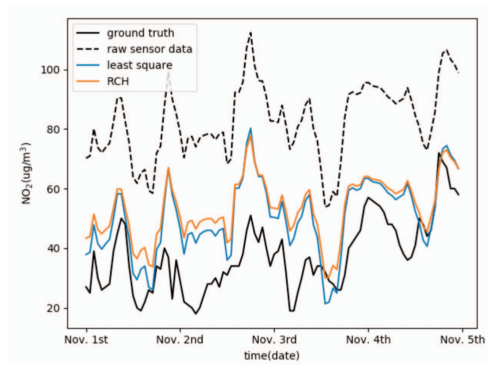

Figure 2: A typical example: RCH got similar calibration effect compared with least-square calibration method

4 CONCLUSION

In this paper we propose RCH, a robust calibration approach for low-cost air quality sensors using historical data of the reference station. RCH is proved to be robust when real-time and nearby reference are unavailable. And RCH have similar performance compared with least-square calibration, a conventional field calibration method with real-time reference. In the future, we will incorporate temporal and spatial prediction method of air pollution data for having more accurate reference to further boost the performance.

ACKNOWLEDGMENTS

This work was funded by National Key Research and Development Project of China 2017YFC0212100 and Shenzhen Science and Technology Program KQTD20170810150821146.

REFERENCES

[1] Matthias Budde, Lin Zhang, and Michael Beigl. 2014. Distributed, low-cost particulate matter sensing: scenarios, challenges, approaches. In *Proceedings of the 1st International Conference on Atmospheric Dust*. 230–236.
[2] Yun Cheng, Xiaoxi He, Zimu Zhou, and Lothar Thiele. 2019. ICT: In-field Calibration Transfer for Air Quality Sensor Deployments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 6.
[3] Andrea Di Antonio, Olalekan A. M. Popoola, Bin Ouyang, John Saffell, and Roderic L. Jones. 2018. Developing a Relative Humidity Correction for Low-Cost Sensors Measuring Ambient Particulate Matter. *Sensors* 18, 9 (2018). https://doi.org/10.3390/s18092790
[4] Balz Maag, Zimu Zhou, and Lothar Thiele. 2018. A survey on sensor calibration in air pollution monitoring deployments. *IEEE Internet of Things Journal* 5, 6 (2018), 4857–4870.
[5] World Health Organization et al. 2018. Ambient air pollution-a major threat to health and climate. *Ambient air pollution* (2018).