Few-shot calibration of low-cost air pollution (PM$_{2.5}$) sensors using meta-learning

Kalpit Yadav[†], Vipul Arora[†], Sonu Kumar Jha[‡], Mohit Kumar[‡], Sachchida Nand Tripathi[‡]

[†]Department of Electrical Engineering, IIT Kanpur

[‡]Department of NAF, IIT Kanpur

{kalpit, vipular, ksonu, mohitk, snt}@iitk.ac.in

Abstract—**Low-cost particulate matter sensors are transforming air quality monitoring because they have lower costs and greater mobility as compared to reference monitors. Calibration of these low-cost sensors requires training data from co-deployed reference monitors. Machine Learning based calibration gives better performance than conventional techniques, but requires a large amount of training data from the sensor, to be calibrated, co-deployed with a reference monitor. In this work, we propose novel transfer learning methods for quick calibration of sensors with minimal co-deployment with reference monitors. Transfer learning utilizes a large amount of data from other sensors along with a limited amount of data from the target sensor. Our extensive experimentation finds the proposed Model-Agnostic-Meta-Learning (MAML) based transfer learning method to be the most effective over other competitive baselines.**

Index Terms—**Air quality, low-cost sensor calibration, few-shot learning, MAML, deep learning**

I. INTRODUCTION

Elevated levels of air pollutants is regarded as the biggest environmental health risk in the whole world. Since traditional air quality monitoring stations have high setup cost and are cumbersome to maintain, administration typically only deploys a few at important locations of a city. As levels of air pollutants depends upon lot of factors and can change significantly from one location to another that are only few kilometers apart, there is an urgent need for higher resolution of data in order to properly assess people's exposure to air pollution.

Fortunately, low-cost sensors offer an affordable solution due to their lower maintenance costs, small size and greater mobility [2]–[4], [8]. These sensors are known to be less accurate than the reference grade monitors, and hence, in-field calibration is required to obtain a reasonable sensor readings. Two settings are commonly used, (i) calibrating the sensors against a co-deployed reference monitor, and (ii) collating the spatio-temporal measurements in a network of sensors [1], [7]. In this paper, we focus on the first setting.

Calibration involves capturing the complex non-linear relationship between the raw sensor values and the true values as measured by the reference monitors. Various statistical methods such as ARIMA [22], [23] and Kalman filtering, and supervised machine learning methods such as linear regression, nearest neighbors and support vector regression have been studied to calibrate low cost sensor readings [5], [6], [17].

Deep learning methods such as fully connected neural network [19], [20], [24], convolutional neural network (CNN) [21], [24], recurrent neural network (RNN) [18], [24] have also been very effective in achieving the state-of-the-art performance in sensor calibration.

In general, deep learning based methods are known to have superior performance compared to conventional machine learning methods in various learning problems [13]. But the former typically require much larger amount of training data compared to the latter [14]. One way to alleviate this problem of data scarcity is transfer learning, which leverages knowledge from abundant data from similar tasks (called as source) and fine-tunes the model over the target task [11], [12]. A state of the art transfer learning technique is model-agnostic meta learning (MAML) [9], [10]. MAML learns an optimal initialization values for the model parameters, so that the model can quickly adapt to the target data. Mostly these advanced deep learning methods address classification problems, while regression brings in its own challenges [15], [16].

In order to address the problem of data scarcity in sensor calibration, various transfer learning and domain adaptation methods have been proposed that attempt to transfer knowledge from a data rich source location to a target location with insufficient data. However, only few studies have been done to transfer the generalised knowledge from multiple source locations to the target location to reduce the risk of negative transfer and achieve meaningful results in the context of sensor calibration.

This paper deals with deep learning based calibration techniques in resource-constrained setting. We wish to minimize the amount of data from a co-located reference monitor needed to prepare the calibration model for a given sensor. We propose a novel MAML based transfer learning method for the regression problem at hand.

II. OUR METHOD

We have multiple low-cost sensors deployed at multiple locations, with one sensor per site. For a sensor, deployed at certain location p, $D_p = (\boldsymbol{x}_p, \boldsymbol{y}_p)$ denote the time series of readings $\boldsymbol{x}_p$ obtained from the low-cost sensor and the corresponding readings $\boldsymbol{y}_p$ obtained from the co-deployed reference monitor, respectively. Both $\boldsymbol{x}_p$ and $\boldsymbol{y}_p$ have same length, denoted by $|D_p|$. We have a set of source locations

P^s, and some target locations P^t. Each target location has a small amount of data, i.e. $|D_{p_i}| >> |D_{p_j}| \forall p_i \in P^s, \forall p_j \in P^t$.

We consider a model, denoted by f_θ with parameters θ that maps observations x to output $\hat{y}$. We use a neural network to implement f_θ.

A. MAML based Sensor Calibration

During meta training, we sample support set and query set from data D_p^s of a source location p. The *support set* is used for learning the parameters of the network, and the *query set* is used for evaluating the performance of the network for that particular location.

B. Proposed Model

Our meta learning framework learns a meta learner from multiple source locations, and uses it to train a base model at the target location, using the limited amount of training data from the target location. The base model is location specific, while the meta learner is generalized for all locations.

We implement a base model as a fully connected deep neural network represented by a function f_{θ_p} with parameters θ_p. To effectively and quickly train the base model for any location p, we take help of a meta learner, which is again implemented as a full connected deep neural network, f_ϕ with parameters ϕ. The architecture of f_ϕ is same as that of f_{θ_p}. In MAML, we use the meta learner to initialize the base model for a new location.

First, we use the source data to train our meta learner. Then, use the limited training data from the target location along with the meta learner to train our base model at a particular target location. The meta model f_ϕ is used to initialize the base learner f_{θ_p} at the target location p.

1) Training the Meta Learner at Source: The meta learner f_ϕ is initialized randomly. We choose a particular source location p_i, the data D_{p_i} is split into a *support set* $D^{\mathcal{S}_{p_i}}$ and a *query set* $D_{p_i}^{\mathcal{Q}}$, which are both mutually exclusive. We use the support set to train the base model, while the query set is used to evaluate the base model and to train the meta learner.

The base model is initialised as ϕ and trained using gradient descent over $D_{p_i}^{\mathcal{S}}$.

$$\theta_{p_i} \leftarrow \theta_{p_i} - \alpha \nabla_{\theta_{p_i}} \mathcal{L}_{D_{p_i}^{\mathcal{S}}}(\theta_{p_i}) \qquad (1)$$

Here, $\alpha \in \mathbb{R}$ is the learning rate. The loss function we use is the mean absolute error, which is defined as

$$\mathcal{L}_D(\theta) = \sum_{(\boldsymbol{x},\boldsymbol{y}) \in D} |f_\theta(\boldsymbol{x}) - \boldsymbol{y}| \qquad (2)$$

The meta learner is trained on the query sets $\{D_{p_i}^{\mathcal{Q}}\}$ over a batch of locations i sampled from source locations. The goal of this training is to optimize ϕ as

$$\phi \leftarrow \phi - \beta \sum_i \nabla_\phi \mathcal{L}_{D_{p_i}^{\mathcal{Q}}}(\theta_{p_i}) \qquad (3)$$

Here, $\beta \in \mathbb{R}$ is the learning rate.

Finally, it is the meta learner only that is saved. The base models at different source locations are discarded.

2) Training the Base Model at the Target Location: To improve the prediction for a target location $p_j \in P^t$ with limited amount of training data, we initialise the network's parameters with ϕ and use Eq.(1) to train θ_{p_j}. The support set here acts as the training data, while the calibration performance is assessed using the query set.

The entire training algorithm of the calibration model is outlined as a pseudo code in Algorithm 1.

Algorithm 1 MAML based sensor calibration

Require: P^s: Set of source locations; P^t: set of target locations.
Require: α, β: step size hyperparameters
1: randomly initialize ϕ
2: **while** not done **do**
3:      Sample a batch of locations from P^s
4:      **for** each location p_i **do**
5:          Sample support datapoints $D_{p_i}^{\mathcal{S}} = \{x_p, y_p\}$ from p_i
6:          Evaluate $\nabla_{\theta_{p_i}} \mathcal{L}_{D_{p_i}^{\mathcal{S}}}(\theta_{p_i})$
7:          Compute adapted parameters with one or more gradient descent steps: $\theta_{p_i} \leftarrow \theta_{p_i} - \alpha \nabla_{\theta_{p_i}} \mathcal{L}_{D_{p_i}^{\mathcal{S}}}(\theta_{p_i})$
8:          Sample query data-points $D_{p_i}^{\mathcal{Q}} = \{x_p, y_p\}$
9:          Evaluate $\mathcal{L}_{D_{p_i}^{\mathcal{Q}}}(\theta_{p_i})$
10:      **end for**
11:      Update $\phi \leftarrow \phi - \beta \sum_i \nabla_\phi \mathcal{L}_{D_{p_i}^{\mathcal{Q}}}(\theta_{p_i})$
12: **end while**
13: **for** each target location $p_i \in P^t$ **do**
14:      Initialise the network with ϕ
15:      Compute the adapted parameters θ_{pi} with few gradient descent using train set
16:      Predict on test set
17: **end for**

III. EXPERIMENTAL RESULTS

A. Dataset Description

We collected co-located data from 15 different locations around Mumbai city for our experiments. The low-cost sensors give four measurements, namely, $PM_{2.5}$, PM_{10}, temperature and humidity, while those from the reference monitor consist of only $PM_{2.5}$ values. The low-cost sensors use Nova fitness sensors, which work on the principal of laser scattering, to measure $PM_{2.5}$. The data was collected from November 1, 2020 to March 5, 2021, with a sampling rate of 1 sample per hour. For our transfer learning experiments, we use 10 locations as source and 5 locations, individually, as the target locations.

From the dataset in each source location, we sample a window of 2 days for the support set and another window of 2 days as the query set. The two sets are mutually exclusive and the query set occurs later in time. For each target city, we sample a window of 3 days as training and validation set, and the next 15 days data for testing.

Loc.	MAE (standard deviation)				RMSE				R^2			
	B_1	B_2	B_3	MAML	B_1	B_2	B_3	MAML	B_1	B_2	B_3	MAML
1	36.4 (29.1)	13.7 (15.2)	9.0 (18.2)	8.0 (6.5)	46.6	16.5	11.2	10.3	-32.4	46.6	75.3	79.9
2	21.8 (31.2)	14.5 (27.9)	12.8 (25.3)	12.1 (17.5)	32.0	24.3	21.7	20.2	0.0	41.4	53.0	55.7
3	18.9 (17.1)	12.8 (14.9)	10.0 (17.2)	8.7 (8.2)	24.7	15.8	14.1	12.0	-56.7	46.9	56.3	68.9
4	21.5 (22.6)	17.1 (19.2)	19.2 (18.3)	15.9 (18.7)	32.3	25.2	26.4	24.5	-90.0	-10.4	-12.1	2.5
5	18.8 (20.9)	14.6 (19.4)	12.6 (21.0)	10.1 (8.2)	24.1	18.8	16.3	12.9	22.1	51.4	64.5	77.0

B. Implementation Details

For the proposed method, we model f_θ as a standard feed-forward neural network. It consists of two hidden layers, each of which has 128 neurons, each with ReLU activation. The output layer has single linear neuron. For training the base model, we use stochastic gradient descent (SGD) optimizer with a learning rate of 10^{-3}, and for training the meta learner, we use Adam optimizer with a learning rate of 10^{-4}.

C. Baselines

We compare our proposed MAML based model with three baselines methods. The first baseline, B_1, uses the limited amount of training data from target location to train the calibration model without any transfer. This is a common training method employed in basic deep learning. The second baseline, B_2, trains the model using training data from a single source location and then transfers the model by fine-tuning it on the target location. This fine-tuning uses Adam optimization and stops when the validation loss stops improving. The third baseline, B_3, trains the model using combined data from all the source locations and then fine-tunes the model on a target location as in B_2.

D. Performance metrics

The performance of a calibration model is assessed by comparing the calibrated $PM_{2.5}$ values against the corresponding readings from the co-located reference monitor. The metrics used for this comparison are mean absolute error (MAE), root mean square error (RMSE) and the coefficient of determination (R^2). We also compute the standard deviation corresponding to the mean absolute error.

E. Results

We implement our method and compare the calibrated $PM_{2.5}$ values using different methods for five target cities. The performance of different calibration methods is shown in Table I. The baseline B_1 performs the worst, showing that deep learning is not effective when the training data is small. The baseline B_2 performs much better as compared to B_1 illustrating the benefits of transfer learning. The performance of transfer learning improves further with the increase in the amount of source data, as supported by the improved results with B_3. However, the proposed method outperforms all others, showing that the proposed MAML based transfer
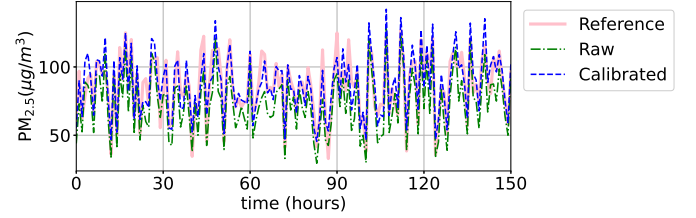

Fig. 1. Example of $PM_{2.5}$ readings from a low-cost sensor (raw), the co-located reference monitor and the proposed calibration model.

learning is better than fine-tuning based transfer learning. While the proposed method shows improvements in MAE, RMSE and R^2 values, we also notice a decline in the standard deviation of absolute error, indicating that the calibrated values follow the reference monitor more closely. Fig. 1 illustrates an example of the calibrated $PM_{2.5}$ values against the raw values from low-cost sensors and readings from reference monitors. We can see the calibrated values follow the reference monitor more closely as compared to the raw sensor readings. Code is available on Github[1].

IV. CONCLUSION AND FUTURE WORK

This paper addresses the problem of quick and efficacious calibration of low cost sensors by minimal co-deployment with reference monitors. It uses deep learning for efficacy and transfer learning for minimal co-deployment. The main contributions of this paper are (i) an efficient solution for calibration of low-cost sensors, and (ii) a model agnostic meta learning based transfer learning method for regression. The proposed method is shown to outperform other learning schemes in the transfer learning setting. The proposed scheme is useful for calibration of static sensors with mobile reference monitors or that of mobile sensors with static reference monitors. In future, with longer deployments, we wish to address sensor drift and maintenance issues. The proposed method is not limited to $PM_{2.5}$ calibration only. We plan to extend our model for calibration of various gas sensors such as those for NO_x, O_3 and CO. We also plan to extend our work to calibration of dense sensor network to monitor the air quality.

[1] https://github.com/madhavlab/2021KalpitBTMT

REFERENCES

[1] Y. You, A. Xu, and T. J. Oechtering. "Belief Function Fusion based Self-calibration for Non-dispersive Infrared Gas Sensor." In 2020 IEEE Sensors, pp. 1-4. IEEE, 2020.

[2] S.H. Shin, and S. Lee. "Multi-channel multi-mode ROIC with embedded calibrations for environmental gas sensors." in IEEE SENSORS, 2017.

[3] J. Dutta, F. Gazi, S. Roy, and C. Chowdhury. "AirSense: Opportunistic crowd-sensing based air quality monitoring system for smart city," In 2016 IEEE SENSORS, pp. 1-3. IEEE, 2016.

[4] S. Esfahani, P. Rollins, J. P. Specht, M. Cole, and J. W. Gardner. "Smart City Battery Operated IoT Based Indoor Air Quality Monitoring System." In 2020 IEEE Sensors, pp. 1-4. IEEE, 2020.

[5] T. Zheng, M. H. Bergin, R. Sutaria, S. N. Tripathi, R. Caldow, and D. E. Carlson, "Gaussian process regression model for dynamically calibrating and surveilling a wireless low-cost particulate matter sensor network in Delhi," Atmos. Meas. Tech., vol. 12, no. 9, pp. 5161–5181, 2019.

[6] R. Sahu, et al., "Robust statistical calibration and characterization of portable low-cost air quality monitoring sensors to quantify real-time O_3 and NO_2 concentrations in diverse environments," Atmos. Meas. Tech. Discuss., no. June, 2020.

[7] J. Hofman, M. E. Nikolaou, T. H. Do, X. Qin, E. Rodrigo, W. Philips, N. Deligiannis, V.P. La Manna, "Mapping air quality in IoT cities: cloud calibration and air quality inference of sensor data," IEEE Sensors, 2020

[8] J. Mitrovics, "Smart sensors for air quality monitoring: Concepts and new developments." In 2016 IEEE SENSORS, pp. 1-2. IEEE, 2016.

[9] C. Finn, P. Abbeel, and S. Levine. "Model-agnostic meta-learning for fast adaptation of deep networks," in ICML, 2017.

[10] Q. Sun, Y. Liu, T. S. Chua, and B. Schiele, "Meta-transfer learning for few-shot learning," in IEEE CVPR, 2018, pp. 403–412.

[11] D.K. Naik and R. J. Mammone. "Meta-neural networks that learn by learning." IEEE International Joint Conference on Neural Networks, 1992.

[12] V. Arora, A. Lahiri, and H. Reetz, "Attribute based shared hidden layers for cross-language knowledge transfer," in IEEE Spoken Language Technology Workshop (SLT), 2016.

[13] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. "Deep learning," MIT press, 2016.

[14] F. Li, R. Fergus, and P. Perona. "One-shot learning of object categories," IEEE Trans. Pattern Anal. Mach. Intell., 28(4):594–611, 2006.

[15] K. Q. Weinberger and G. Tesauro, "Metric learning for kernel regression," J. Mach. Learn. Res., vol. 2, pp. 612–619, 2007.

[16] S. Lathuilière, P. Mesejo, X. Alameda-Pineda, and R. Horaud. "A comprehensive analysis of deep regression." IEEE transactions on pattern analysis and machine intelligence, vol. 42, no. 9: 2065-2081, 2019.

[17] D. Holstius, A. Pillarisetti, K. R. Smith, and E. Seto. "Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California." Atmospheric Measurement Techniques, vol. 7: 1121-1131, 2014.

[18] Han, P.; Mei, H.; Liu, D.; Zeng, N.; Tang, X.; Wang, Y.; Pan, Y. Calibrations of Low-Cost Air Pollution Monitoring Sensors for CO, NO2, O3, and SO2. Sensors 2021, 21, 256.

[19] L. Spinelle, M. Gerboles, M. G. Villani, M. Aleixandre, and F. Bonavitacola, "Field calibration of a cluster of low-cost available sensors for air quality monitoring. part a: Ozone and nitrogen dioxide," Sensors and Actuators B: Chemical 215:249–257, 2015.

[20] L. Spinelle, M. Gerboles, M. G. Villani, M. Aleixandre, and F. Bonavitacola, "Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. part b: NO, CO and CO2," Sensors and Actuators B: Chemical 238:706–715, 2017.

[21] H. Yu, Q, Li, Y.-A. Geng, Y. Zhang, Z. Wei, "AirNet: A Calibration Model for Low-Cost Air Monitoring Sensors Using Dual Sequence Encoder Networks," in AAAI, 2020

[22] S. Munir, M. Mayfield, "Application of Density Plots and Time Series Modelling to the Analysis of Nitrogen Dioxides Measured by Low-Cost and Reference Sensors in Urban Areas," Nitrogen, 2, 167–195, 2021.

[23] X. Yan and L. Shuangting "Time Series and Multiple Linear Regression Calibration Model for CO Monitoring Data," in E3S Web of Conferences, vol. 214, p. 03039. EDP Sciences, 2020.

[24] E. Bagkis, T. Kassandros, M. Karteris, A. Karteris, K. Karatzas "Analyzing and Improving the Performance of a Particulate Matter Low Cost Air Quality Monitoring Device". Atmosphere 2021, 12, 251.