Improving Air Pollution Forecast with Ubiquitous Mobile Sensor Network

Ke Hu[*1], Ashfaqur Rahman[†2], Vijay Sivaraman[*3] and Pradeep Ray[*4]

[*]Asia Pacific ubiquitous Healthcare research Centre (APuHC), UNSW Australia

[†]Autonomous Systems Program, CSIRO Computational Informatics, Hobart Tasmania, Australia

[1]ke.hu@student.unsw.edu.au, [2]ashfaqur.rahman@csiro.au, [3]vijay@unsw.edu.au, [4]p.ray@unsw.edu.au

Abstract—**Recently, high spatial resolution is increasingly needed in air pollution forecast by governments and climatologists. However, data from fixed monitoring sites cannot meet these requirements. In this paper, we present a hybrid spatio-temporal model to improve the air quality forecast capacity with fine grid data from ubiquitous mobile sensor network.**

Keywords—*air quality; ubiquitous sensor network; mobile apps;*

I. Introduction

Air pollution has became a worldwide problem during the last two decades. The World Health Organization (WHO) estimates that millions of people died every year due to air pollution. [1] Nowadays, continuously monitoring and forecasting air quality are becoming increasingly important for controlling and solving the air pollution problem.

Some government agencies do supply air pollution forecast services to public [2], [3], while many research groups also attempt to develop new models to predict air pollution exposure. For example, the authors of [4] developed a new fuzzy time series model which contains six steps for forecasting pollution concentrations. Weather forecast and meteorological situation data was included in their prediction. However, they did not take spatial information in consideration, therefore they could only predict air pollution concentrations around the weather stations. In [5], researchers used a two-step methodology which firstly used neural network(NN) to forecast the pollution concentrations around the monitoring stations, and then used spatial stochastic simulation method to evaluate the concentrations in the entire area. The air quality data they used was from 12 monitoring stations which were separated few kilometres from each other at least within the greater Lisbon area. The forecast results may be biased with such a low density air pollution data source.

In our study, we introduce a hybrid spatio-temporal model based on data from ubiquitous mobile sensor network to improve the air pollution forecast capacity. Our contribution is to build a model to improve the air quality forecast performance both in spatial and temporal based on new ubiquitous mobiles sensor network. The simulation results show that with our model and data from ubiquitous sensor network, air pollution prediction capacity and accuracy can be highly improved.

The rest of this paper is organized as follows: Section II describes our hybrid spatio-temporal model. In Section III we present simulation results based on short-term air pollution data which is from mobile sensor network and fixed monitoring sites, and the paper concludes in Section IV.

II. Hybrid Model Description

Our model is developed based on a two-step hybrid methodology as describes in Table I.

TABLE I. Hybrid model

1.Spatial interpolation model
1.1. Separate spatial map into grid cells.
1.2. Use IDW (Inverse distance weighting) interpolation model, along with air pollution data from fixed stations and mobile sensor network to calculate the concentration of every grid cell (Assume the center point of each grid cell to be its location).
2.Temporal forecasting model
2.1. Use neural network (NN) to do the temporal forecasting of a pollutant's concentration in each grid cell with past air pollution concentration values.

A. Spatial interpolation model

Let $Y(s,t)$ donates the air pollution concentration at location s and time t. Assume we separate the forecast map into N grid cells, s can be any location in the center of each grid cell. Following the IDW interpolation model, then the pollution concentration in every location s in a certain time T based on sample $y_i = y(x_i)$ for $i = 0, 1, ..., M$ can be as follows:

$$Y(s,T) = \sum_{i=0}^{M} \frac{\omega_i(X)y_i}{\sum_{j=0}^{M} \omega_j(X)}, \quad (1)$$

where

$$\omega_i(X) = \frac{1}{d(s,x_i)^2}, \quad (2)$$

B. Temporal forecasting model

After we get $Y(s,t)$ in a certain period of time t, we then can use neural network to train these data to forecast the temporal air pollution concentrations. Let us consider use $Y(S,t_p)$ which can present the pollutant's concentrations at location S and time p as inputs and prediction result $Y(S,t_q)$ would be as follows:

$$Y(S,t_q) = \sum_{i=p}^{q-1} Y(S,t_i)\omega_i, \quad (3)$$

where ω_i is the weight of every input and q is the time of prediction result.
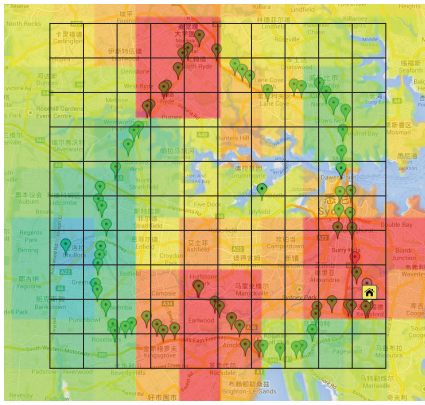
Fig. 1. Grid map in Sydney (Different colors indicate various levels of pollutant's concentrations. The concentrations increase from cool colors to warm colors.

Then we can train the model and perform a cross-validation experiment to create an air pollution forecasting of every $Y(s, t_{p+q})$.

TABLE II.   AIR POLLUTION DATA (CARBON MONOXIDE) SUMMARY

Data source	Min.(ppm)	Mean(ppm)	Max.(ppm)
Mobile sensor network	0.00700	4.214793	47.540062
Fixed sites	0.3	0.3	0.3

TABLE III.   RESULTS OBTAINED BY DIFFERENT TIME PERIODS AHEAD FORECASTING

Time (minutes)	Mean absolute error(ppm)	Accuracy
5	1.084400	49.0894%
6	1.065777	48.8992%
7	1.111915	52.4867%
8	1.146594	53.9134%

TABLE IV.   RESULTS COMPARE FROM DIFFERENT DATA SOURCE

Data source	Forecasting result(ppm)	Accuracy
Ubiquitous sensor network	11.849259	51.8843%
Fixed stations	0.3	3.75%

III.   SIMULATION RESULTS

We used a mobile sensor network system named Haze-watch [6] to upload Carbon Monoxide (CO) concentration data, as well as data from fixed monitoring sites in Sydney. The system architecture contains four parts: **Sensor part**: Air pollution sensors will collect data and transfer it to mobile phone via bluetooth. **Mobile part**: Mobile phone will upload air pollution data along with the location and time stamp. **Server part**: Server can store data, generate GIS map, and respond the enquiries from users. **User part**: Users can use devices like mobile phone, pad, laptop or PC to run applications to check the air pollution concentrations and other information. Data is updated per 10 seconds from sensor network system following a route as shown in Fig. 1 and hourly from two fixed monitoring stations. Air pollution data we use in this simulation is collected within 2 hours (11:00 to 13:00) and summarized in Table II. CO values from sensor network range from 0 (ppm) to 47 (ppm) while data from fixed sites remains low (0.3 ppm).

After data has been collected, we separate the map of Sydney into 100 grid cells as shown in Fig. 1. Then we use

IDW to calculate the CO concentration in each grid cell during 2 hours period with 1 minute gap (As the air pollution data is updated per 10 seconds, we use an average value to present one minute concentration), then apply few minutes (5, 6, 7, 8) data ahead to train the NN model to forecast the pollution concentration in next minute with 1 hour data. We validate the other 1 hour data after the model has been trained. If applying the spatial model only, we cannot get the prediction results; Also, only temporal model being used in the forecasting can get no other results than few prediction outcomes around the fixed stations. As a result, we only perform simulations with our hybrid model.

The simulation results are shown in Table III. We can see that with our model, the forecast accuracy can achieve 53.9% with a small group of data. And the forecasting accuracy will increase with the rising of previous data that be used in the prediction. To compare the forecasting accuracy between different data source, we choose one location in the center of Sydney, and use one commercial air pollution monitoring device (Honeywell GasAlertmicro5) to get the real-time CO concentration value at 12:50, which is 8 (ppm). Then we use our hybrid model with data from sensor network and fixed monitoring stations to do the forecast. As we can see from Table IV, with data from sensor network, the prediction accuracy (51.8843%) can be much higher than with data from fixed stations (3.75%).

IV.   CONCLUSIONS AND FUTURE WORK

In this paper we have presented a hybrid spatio-temporal model to improve the air pollution forecast performance with data from ubiquitous sensor network, and the simulation results show that forecasting accuracy and capacity can be highly improved by our model. We know that there are many variables like meteorological information are not considered in this paper. In future work, we can use meteorological sensors like portable weather station to collect data along with the air pollution data to increase the prediction accuracy, and use method in [7] to involve personal activity information in forecasting to personalize the results. More data mining approaches also can be explored in our model.

REFERENCES

[1] World-Health-Organization, "Air Pollution." [Online]. Available: http://www.who.int/topics/air_pollution/en/

[2] AIRNow, "Air Quality Index forecast." [Online]. Available: http://www.airnow.gov/

[3] The-Weather-Channel, "Air Quality Forecast." [Online]. Available: http://www.weather.com/activities/health/airquality/

[4] D. Domaska and M. Wojtylak, "Application of fuzzy time series models for forecasting pollution concentrations," *Expert Systems with Applications*, vol. 39, pp. 7673 – 7679, 2012.

[5] A. Russo and A. Soares, "Hybrid model for urban air pollution forecasting: A stochastic spatio-temporal approach," *Mathematical Geosciences*, vol. 46, pp. 75–93, 2014.

[6] V. Sivaraman, J. Carrapetta, K. Hu, and B. Gallego Luxan, "Hazewatch: A participatory sensor system for monitoring air pollution in sydney," in *IEEE SenseApp (co-located with IEEE LCN)*, Sydney, Australia, Oct. 2013.

[7] K. Hu, Y. Wang, A. Rahman, and V. Sivaraman, "Personalising pollution exposure estimates using wearable activity sensors," in *IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, Singapore, Apr. 2014.