Improving Air Pollution Forecast with Ubiquitous Mobile Sensor Network

Ke Hu, Vijay Sivaraman, Pradeep Ray
(Asia Pacific ubiquitous Healthcare research Centre (APuHC))
& Ashfaqur Rahman
(Autonomous Systems Program, CSIRO Computational Informatics)

IEEE ICUFN, 10 Jul 2014

Air pollution: effects

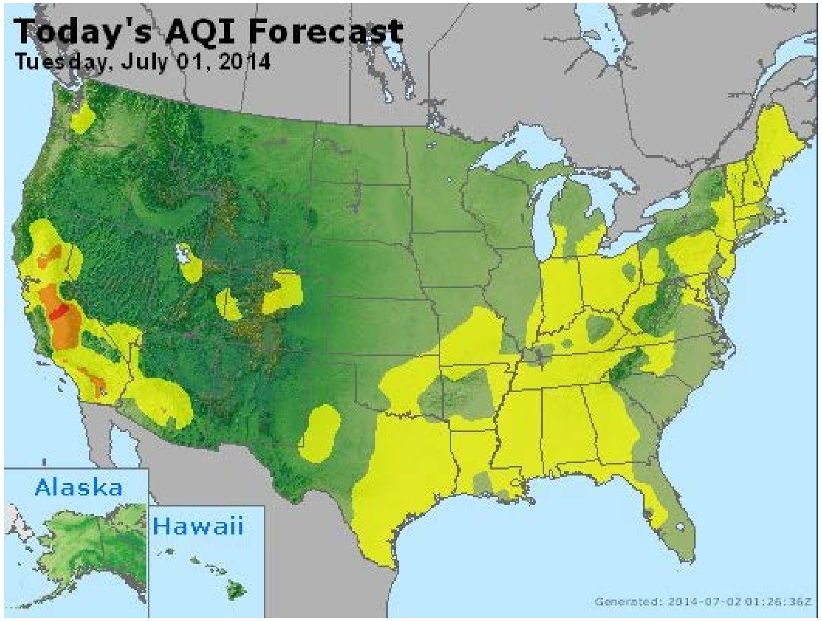
- Air pollution killed **seven** million people in 2012
  - More than Aids, diabetes and road accidents combined
- Air pollution causes 1 in 8 deaths worldwide
- Air pollution becomes the world's **largest** environmental health risk

UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA

Images From: http://environment.nationalgeographic.com

Air pollution forecast

- Why forecast air pollution?
  - Guide citizens to react properly to air pollution (Wear face masks, avoid heavily polluted area, etc.)
  - Support air pollution control
- Forecast status and limitation
  - Some government agencies do supply air pollution forecast services to public
  - Air quality data is from fixed sites which are in a **low density**.



Images From: http://www.airnow.gov/

Our proposal :

- A hybrid spatio-temporal model
  - Based on data from ubiquitous mobile sensor network
  - Improve the air pollution forecast capacity

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.

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 $i = 0, 1, \ldots, M$ can be as follows:

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

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

Temporal forecasting model

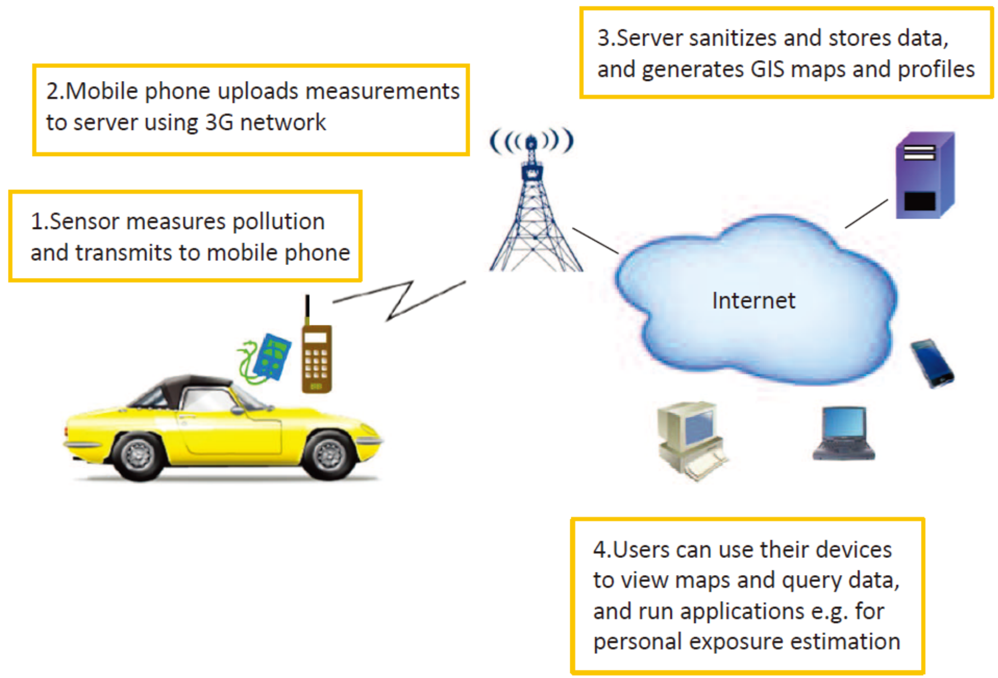
- After we get Y (s, t) in a certain period of time t, we then can use neural network to train these data to forecast the temporal air pollution concentrations. Let us consider use Y (S, t_p) which can present the pollutant's concentrations at location S and time p as inputs and prediction result Y (S,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

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

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

Hazewatch system

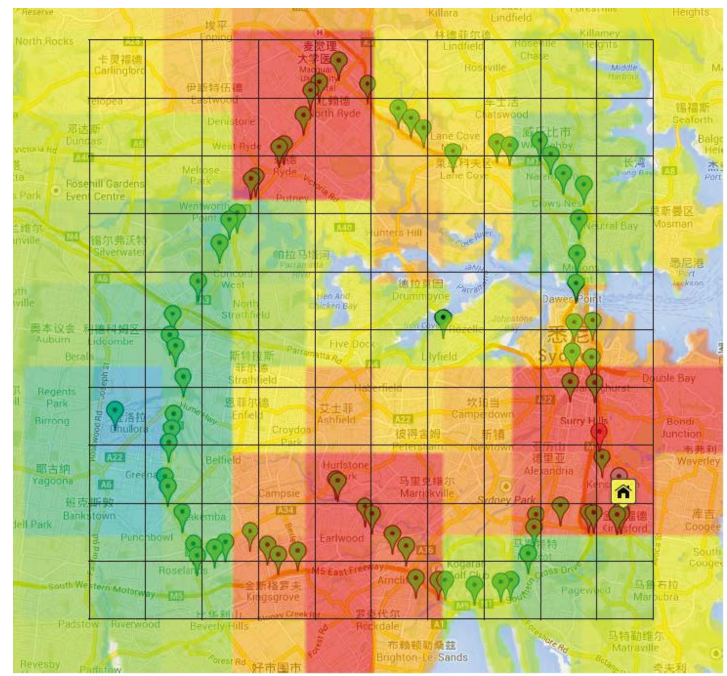
- Sensor part: Air pollution sensors
- Mobile part: Mobile phone
- Server part: Cloud server
- User part: Applications



Simulation preparation:

- Carbon Monoxide data
  - Data is updated per 10 seconds from sensor network system and hourly from two fixed monitoring stations
  - Collected within 2 hours
- Separate the map of Sydney into 100 grid cells

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



UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA

Simulation

- Spatial interpolation model
  - Calculate the CO concentration in each grid cell during 2 hours period with 1 minute gap
- Temporal forecasting model
  - 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.
  - Validate the other 1 hour data after the model has been trained.

Simulation results

- Forecast accuracy can achieve 53.9% with a small group of data.
- Forecasting accuracy will increase with the rising of previous data that be used in the prediction.

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%

UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA

Simulation results

- Compare the forecasting accuracy:
  - Use one commercial air pollution monitoring device (Honeywell GasAlertmicro5) to get the real-time CO concentration value at a certain location.(8ppm)
  - Use our hybrid model with data from sensor network and fixed monitoring stations to do the forecast

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

Conclusion

- We presented a hybrid spatio-temporal model to improve the air pollution forecast performance with data from ubiquitous sensor network

- Our simulation results show that forecasting accuracy and capacity can be highly improved by our model

- Future work
  - Use meteorological sensors like portable weather station to collect data along with the air pollution data to increase the prediction accuracy
  - More data mining approaches also can be explored in our model

Thank you!

Email: jasonhu.au@gmail.com