Air Pollution Exposure Estimation and Finding Association with Human Activity using Wearable Sensor Network

Ke Hu, Timothy Davison, Vijay Sivaraman

(School of Electrical Eng. & Telecommunications ,UNSW, Australia)

& Ashfaqur Rahman

(Autonomous Systems Program, CSIRO Computational Informatics)

Workshop on Machine Learning for Sensory Data Analysis

02 Dec 2014

Air pollution

- WHO reports air pollution kills about 7 million people a year
- World's largest single environmental health risk

- Environmental issues
  - Climate change
  - Global warming

Images From UNPE: http://www.unep.org/

Monitor air pollution

Fixed sites monitoring network

- Usually operated by government agencies
- A range of pollutants
- High cost and space required
- Sparse density

Participatory sensor network

- Idea: Crowdsourcing
- Sensor nodes
- Mobile network

Motivation:

Are general pollution concentrations enough?

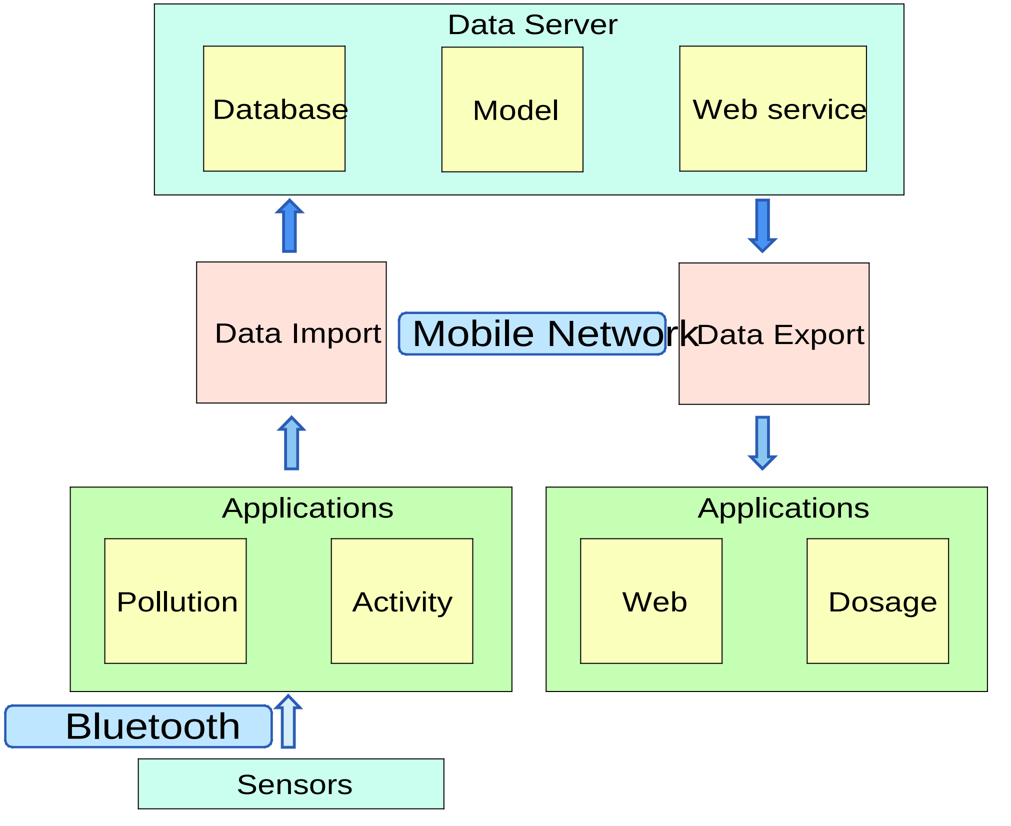
- Individuals are more concerned about their personal air pollution dosage

Include activity information to personalize the air pollution influence

What is the association between activity modes, locations and inhalation dosage?

Our proposal :

- A "Crowd source" system to estimate personal air pollution inhalation dosage
  - Data from users
  - Both air pollution data and activity data is collected
  - Inhalation dosage data can be displayed on our novel mobile application
  - Users with wearable activity sensors, and those without, can all benefit from our application.



Sensor selection

- Air pollution sensor
  - Node: Plug-in modules mode; Measures various pollutants;
- Activity sensors (Energy expenditure)
  - **Fitbit activity wristband**: Sync the device via Bluetooth
  - Jawbone UP: Needed to be connected to the audio jack

Dosage Calculation algorithms

Estimate Energy Expenditure

Calculate Resting Metabolic Rate (RMR)	Calculate Energy Expenditure (EE)

$$RMR = (0.166) * [a + b * (BM) + e]$$

$$EE = MET * RMR$$

Calculate Inhalation Dosage (ID)

Calculate Oxygen Uptake Rate (VO2)	Calculate Ventilation Rate (VR)

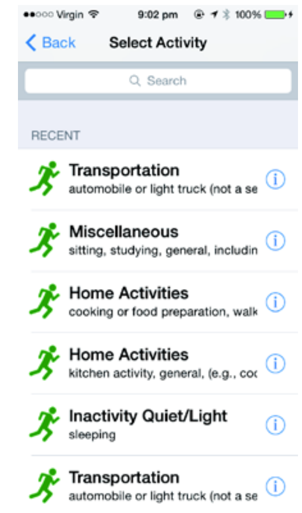
$$VO_2 = ECF * EE$$

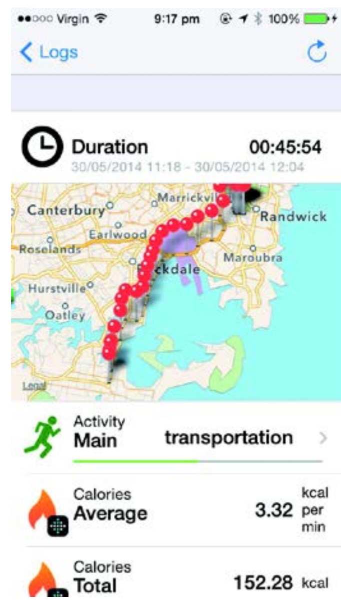
$$VR = BM * e^{c+d*\ln \frac{VO_2}{BM}}$$

$$ID = VR * PC$$

Application

- Age, body mass and gender should be set or sync from Fitbit server
- Records location periodically
- Fetches pollution estimate from our air pollution server
  - User need not carry air pollution hardware
- Activity data:
  - Manually selected
  - Fitbit server
- Displays:
  - Plot map
  - Calories burned
  - Air pollution concentrations
  - Personal dosage





Trial design

- Carbon Monoxide (CO) is selected as the pollutant in this study
  - Air pollution sensors (per 5 seconds)
  - Government fixed monitoring sites. (Hourly)

- One participant
  - Lives in North Ryde and works in UNSW Australia
  - Wears the Fitbit sensor and carries air pollution sensor during a 24 hours period
  - Changes activity modes manually

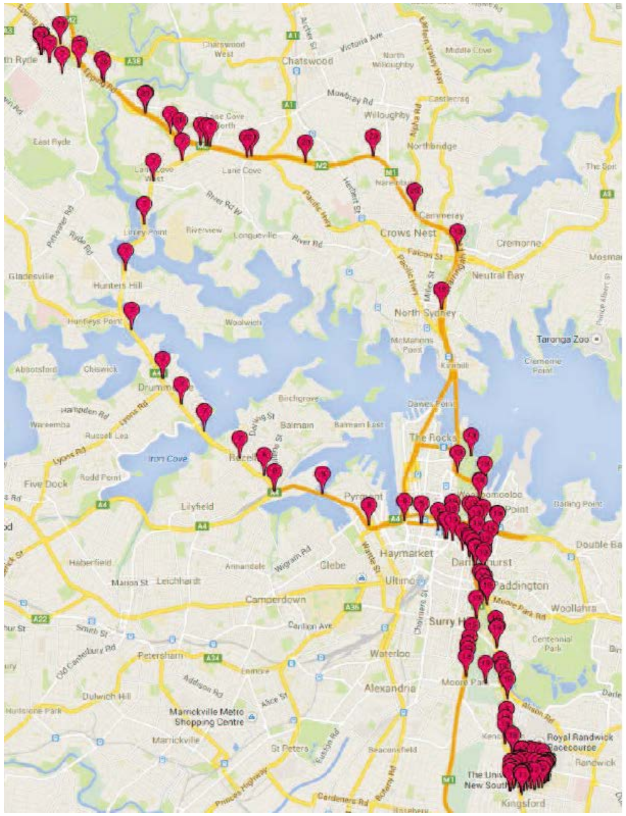
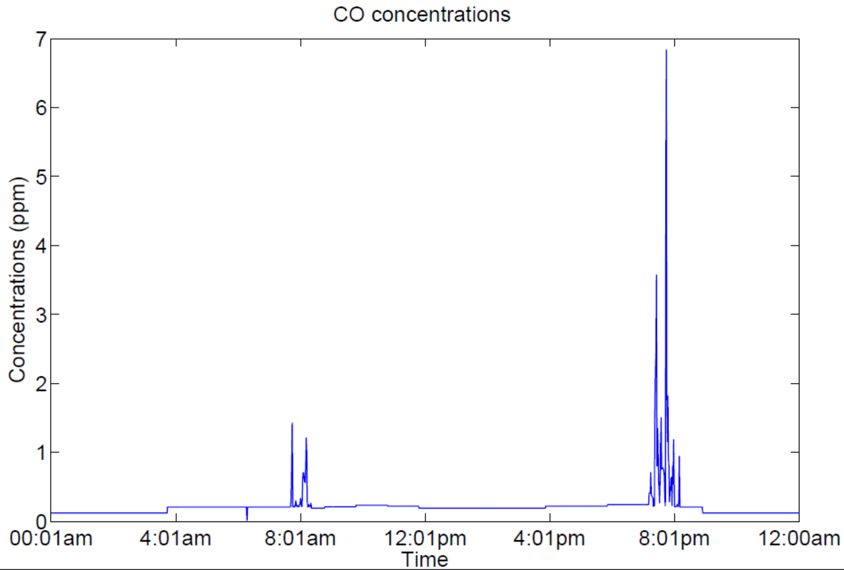


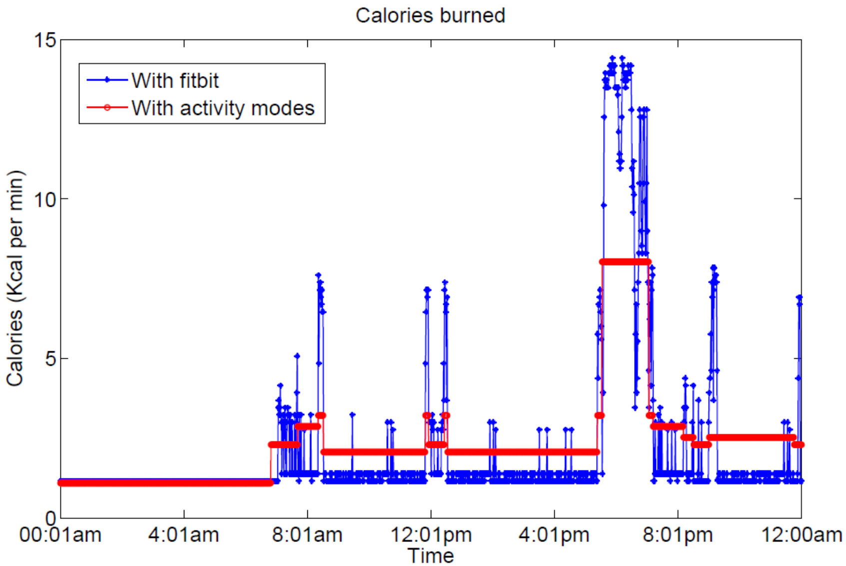
Table 2: Participant general information

Gender	Body mass (kg)	Age	Stature(cm)
Male	67.8	30	180

Result:

- CO concentrations
- Calories burned information:
  - Data from MET is flat while engaging in one activity mode
  - Data is assembled in few time slot

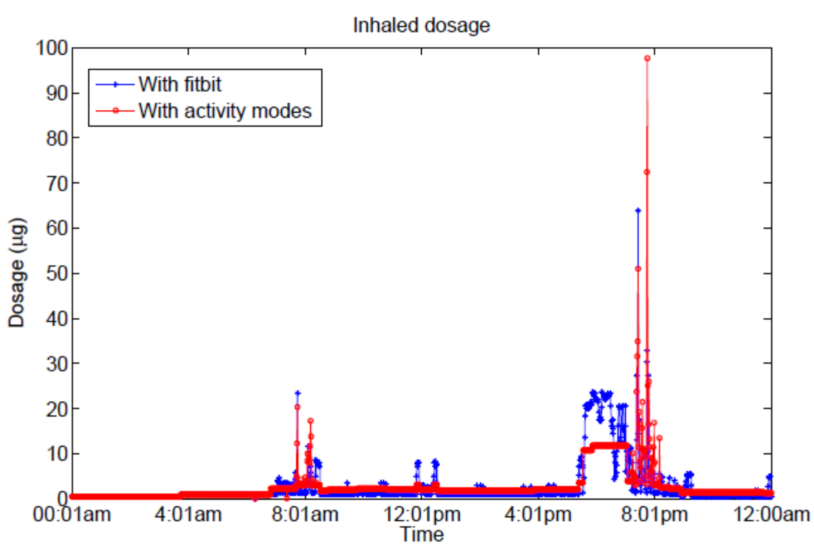


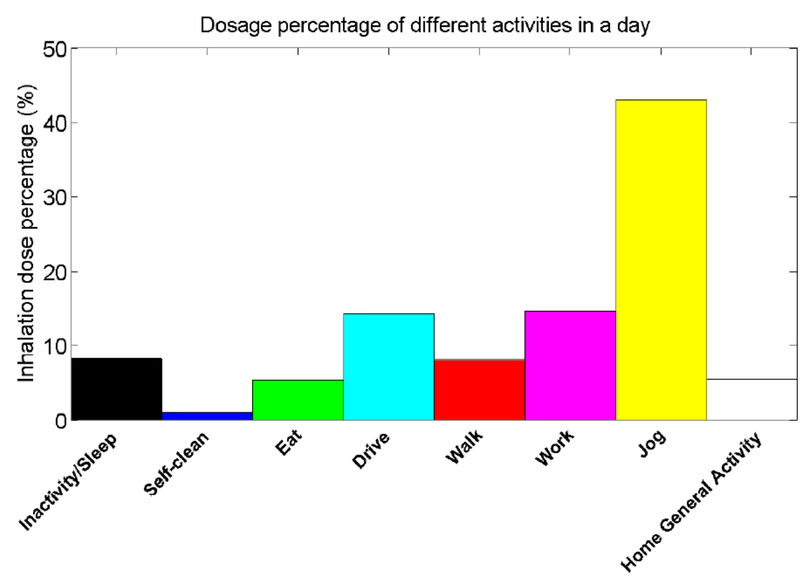


	Time	Activity modes	Latitude	Longitude	Calories (Kcal per min)	CO Concentrations (ppm)	Dosage (μg)
Non-Fitbit	19:22	Drive	-33.8880822904876	151.219555087863	2.866571	1.6621	23.74718802
With Fitbit	19:22	Drive	-33.8880822904876	151.219555087863	3.227280	1.6621	27.36695547

Result:

- Inhaled dosage:
  - Dosage data with or without Fitbit correlate with each other well
  - Distinct during jogging
- Dosage percentage of different activities :
  - Doing sports or fitness may not be as healthy as people think
  - Cannot be confirmed as it is unclear what level of exposure leads to health risk.





Data evaluation

- What is the association between activity modes, location and dosage?

Applied K-means method to cluster the whole data points into three groups based on the dosage levels.

Applied several classification techniques to classify the data in terms of activity modes, locations and dosage levels

	range (μg per min)	No. of instances
Low	4.23 - 0.57	1261
Medium	13.69 - 4.31	101
High	63.85 - 14.58	78

These techniques were running in an open source data mining software WEKA

Data evaluation

- The best accuracy achieved in this work is 94.898% with JRip and J48 algorithms.
- Algorithm LibSVM has the lowest mean absolute error rate and relative absolute error
- The longest training time is taken by MLP algorithm

Classifier	TP rate	FP rate	Precision	Recall	F-measure	ROC area
ZeroR	0.861	0.861	0.742	0.861	0.797	0.500
Naive Bayes	0.912	0.444	0.919	0.912	0.882	0.969
BayesNet	0.916	**0.119**	0.925	0.916	0.920	**0.970**
LibSVM	0.941	0.254	0.941	0.941	0.933	0.843
MLP	0.941	0.254	0.941	0.941	0.933	0.969
JRip	**0.949**	0.129	**0.946**	**0.949**	**0.946**	0.916
J48	**0.949**	0.129	**0.946**	**0.949**	**0.946**	0.965

Classifier	Accuracy	Mean absolute error	Root mean squared error	Relative absolute error	Root relative absolute error	Kappa statistic	Time (Seconds)
ZeroR	86.1224%	0.1550	0.2881	100.0000%	100.0000%	0.0000	**0.00**
Naive Bayes	91.2245%	0.0643	0.2279	41.5029%	79.0845%	0.5448	0.02
BayesNet	91.6327%	0.0545	0.2194	35.1307%	76.1476%	0.6835	0.03
LibSVM	94.0816%	**0.0395**	0.1986	**25.4527%**	68.9389%	0.7288	0.27
MLP	94.0816%	0.0540	0.1801	34.8591%	62.4953%	0.7288	1.59
JRip	**94.8980%**	0.0500	0.1760	32.2426%	61.0965%	**0.7851**	0.08
J48	**94.8980%**	0.0471	**0.1748**	30.3831%	**60.6660%**	**0.7851**	0.03

Data evaluation

- Location will not effect dosage levels in the same activity mode except driving
- Dosage level of sleep, self-clean, eating, working, and home general activity is low
- Dosage level of walking and jogging is medium and high respectively

```
Activity = Inactivity/Sleep: Low (409.0)
Activity = Self-clean: Low (37.0/4.0)
Activity = Eat: Low (91.0/3.0)
Activity = Drive
| Latitude <= -33.877125
| | Longitude <= 151.220242
| | | Longitude <= 151.217461: Medium (14.0/1.0)
| | | Longitude > 151.217461
| | | | Latitude <= -33.886577
| | | | | Longitude <= 151.219215: Low (2.0)
| | | | | Longitude > 151.219215: High (2.0)
| | | | Latitude > -33.886577
| | | | | Longitude <= 151.218538: High (2.0)
| | | | | Longitude > 151.218538: Medium (5.0/2.0)
| | Longitude > 151.220242: Low (20.0/2.0)
| Latitude > -33.877125: Low (54.0/7.0)
Activity = Walk: Medium (39.0/1.0)
Activity = Work: Low (489.0)
Activity = Jog: High (91.0/21.0)
Activity = Home General activity: Low (185.0/12.0)
```

Conclusions

- We presented a novel mobile application that estimates personal air pollution dosage using human energy expenditure and other personal data from wearable activity sensor devices.

- With our field trials data in Sydney, we summarize that:
  - Dosage during sleeping, eating, working in a campus and doing general home activity is low; people will inhale more while working out, walking or driving outdoors.
  - The performance of J48 classier is the best, achieving nearly 94% accuracy within 0.03 seconds.

- Future work:
  - Release our application and gain more data to analyze the human daily dosage, making the classification result more convincing
  - Estimate the dosage based on the predictors instead of just classifying

UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA