Increasing the precision of mobile sensing systems through super-sampling

RJ Honicky[†], Eric A. Brewer[†], John F. Canny[†], Ronald C. Cohen[‡]
[†] Department of Computer Science, UC Berkeley
Email: {honicky,brewer,jfc}@cs.berkeley.edu
[‡] Department of Chemistry, UC Berkeley
Email: rccohen@berkeley.edu

Abstract

Sensors integrated into mobile phones have the advantage of mobility, co-location with people, pre-built network and power infrastructure, and potentially, ubiquity. These characteristics, however, also present significant challenges. Mobility means non-uniform sampling in space, and also constrains the size and weight of the sensors. In this paper, we focus on non-uniform sampling, and imprecision. We investigate the question, "Assuming well calibrated sensors, what precision can we expect from a network of sensors embedded in location aware cell phones?" We briefly describe some results that suggest that a Gaussian process based model is appropriate.

I. Introduction

With increased public focus on environmental conditions and increasing industrialization of developing countries, the need for environmental monitoring has increased significantly. Current air pollution monitoring systems typically consist of highly sensitive, bulky equipment placed in a few strategic locations. These systems, such as the California Air Resource Board (CARB) monitoring system mostly monitor ambient levels over large geographic areas [1]. Not only do systems like CARB have very coarse granularity, but they also only measure the human and environmental health impacts of pollution indirectly.

The Networked Suite of Mobile Atmospheric Real-Time Sensors (N-SMARTS) project [2] aims to radically improve the geographic coverage and granularity of environmental monitoring by integrating pollution (and other environmental) sensors into location-aware mobile phones. Our current sensor devices connect to the phone via Bluetooth, and will eventually fit into a modified battery pack, for tight ergonomic integration. Sensors integrated into mobile phones have the advantage of mobility, co-location with people, pre-built network and power infrastructure, and potentially, ubiquity.

These characteristics, however, also present significant challenges. Mobility means non-uniform sampling in space, and also constrains the size and weight of the sensors. Although co-location with people means that samples will often be taken near a particular person, hence providing a good approximation of a person's exposure to pollution, co-location also means that a person's behavior (putting their phone in their pockets, riding in cars, remaining indoors vs. outdoors) will impact the readings of the sensors. Tracking a person's location also has enormous privacy implications. Ubiquity implies low cost and, coupled with size constraints, low-precision sensors. Embedding sensors into a ubiquitous device also implies a passive sensing model, in which the user can not be expected to perform any action to sense the environment, nor can they be expected to calibrate or otherwise maintain the sensor.

In this paper, we focus on a small piece of this puzzle: non-uniform sampling, and imprecision. We investigate the question, "Assuming well calibrated sensors, what precision can we expect from a network of sensors embedded in location aware cell phones?" We make a case for using a Gaussian Process noise model and show some early empirical and simulation results.

A. Problem formulation

Fundamentally, we are interested in measuring and characterizing the environment using sensors embedded in location-aware mobile phones. For the sake of concreteness, in this paper we focus on carbon monoxide, but we believe that these results will extend to many other environmental factors, including other gaseous pollutants, aerosol pollutants, radiation and network signal strength.

Since we are interested in modeling the environment as people experience it over time, we use a model with two spacial dimensions (people basically move two dimensionally), and a temporal dimension.

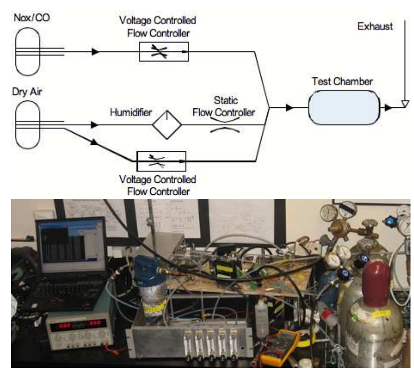


Fig. 1. The test chamber allows precise control of the concentration of toxic gases and fast response, which allows precise calibration and characterization of the sensors.

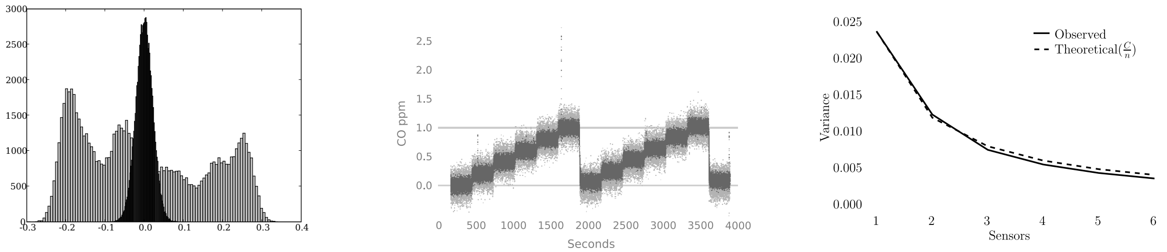
B. Data

In order to understand pollution sensors in greater detail, we have designed a series controlled laboratory experiments. To characterize the CO sensors we are using, we use two electronically controlled mass flow controllers, one attached to pure air, the other attached to 100ppm CO air. The output of the flow controllers is then pumped into a cylindrical chamber that contain six sensor and associated electronics. Finally, the gas is injected into the laboratory's exhaust system (see Figure 1). This setup allows us to precisely control the concentration and rate of flow of CO in the sampling chamber. The sensors and flow controller are monitored and controlled using a NI USB-6218 data acquisition module from National Instruments attached to a laptop.

II. A Gaussian noise model

Sensor noise is often well modeled with a Gaussian distribution. One reason for this is that Gaussian noise turns out to be a good model for a wide range of physical phenomenon, including the thermal noise in electronics.

The CO sensor that we use produces a very faint signal, which makes it vulnerable to ambient noise (e.g. the sensors receive and amplify this noise over the air), including AC power hum. Figure 2(a) shows the noise deviation from the mean of readings from the sensor before and



(a) Senor noise before and after filtering with a 60Hz (and harmonics) notch filter

(b) Sensor readings of concentration of CO(ppm) vs. time. Light dots show the readings from a single sensor. Dark dots show the average of six sensors.

(c) Variance of the average signal from a set of sensors (in ppm) vs. the number of sensors in the set. $\frac{C}{n}$ is show for reference.

Fig. 2. Empirical results with our CO sensors and test chamber.

after the 60Hz hum and its harmonics were removed using notch filters. The filtered noise is Gaussian, providing some empirical justification to assume a Gaussian noise process.

III. Empirical results

As the density of sensors at a given location increases, we can increase our precision by super-sampling, and averaging. For sensors with Gaussian noise (which our CO sensors exhibit) sampling in the same location, we expect the variance of the signal to be $\frac{C}{n}$ if we average the signals from n sensors with noise variance C. Note that when the noise is not Gaussian, the noise power will still decrease, but at a slower rate.

In Figure 2(b), we a experiment with six sensors in a chamber in which we can control the concentration of CO. In this case, we stepped the concentration of CO by 0.2ppm increments over an hour, and observed the response of the sensors. The light dots show the response of one sensor, and the dark dots show the averaged response of six sensors. Clearly the noise variance has decreased. Figure 2(c) show the variance of the signal versus the number of sensors averaged. The empirical results match the theoretical results closely!

IV. Gaussian processes

Using Gaussian process regression (GPR), we can also increase the precision of the system even when samples are not in the same location in space-time (a more realistic situation). The closer the samples are to one another, the greater the increase in the precision.

We should note that a GPR is appropriate not only because the *sensor noise* is Gaussian, but because process by which concentrations of gas mix and vary is also often modeled as Gaussian [3]. Modeled this way, we have the sum of two Gaussians, which is itself a Gaussian. More complex models might include inference of prevailing winds as well, but it remains to be seen if these complications are in fact necessary.

Gaussian process regression is a kernel method, and as such, shares many similarities with other kernel methods such as support vector machines (SVM). It is beyond the scope of this paper to describe the mathematics of GPR. Depending on the kernel, GPR can be as computationally efficient as SVM [4].

V. Learning curves

The amount that the precision of the system increases depends on the density of sampling. As the density of sampling increases, so does the precision.

To quantify this increase in precision for a given algorithm, it is typical to consider the "learning curve" of the algorithm. The learning curve shows the deviation of the true values of samples from the inferred function as the number of training examples increases for a given area. Sollich [5] provides some reasonably tight analytical bounds on the learning curves for GPR. In the future we will present an analysis of the learning curves under various model assumptions.

In Figure 3, we see simulation results in which the variance of the signal at a point decreases when nearby sensor's readings are also taken into account. In this simulation, we use a standard radial basis kernel, and the sensors are uniformly distributed within twice the scale of the kernel. This means that many of the points will be relatively far away from the point of interest, and will not contribute significantly to reducing the variance. Nonetheless, we can see that as the density near the point of interest increases, the variance decreases.

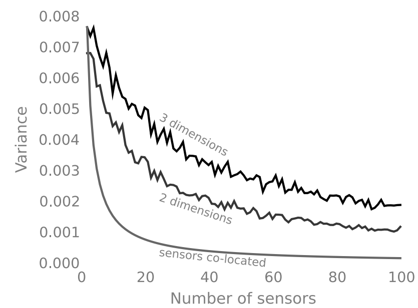


Fig. 3. Simulation of signal variance at a point when samples from different nearby sensors are also utilized vs. the number of nearby sensors. Variance is shown for two dimensional and three dimensional coordinates. For comparison, the variance is show for the case in which all of the sensors sample at the same point in space, as in Figure 2(c).

VI. Future work

This paper begins to explore one way in which mobility in sensors can be exploited to increase the usefulness and (in this case) precision of the sensing system. Although it examines super-sampling under (mostly) ideal situations, many questions remain to be answered. How does miscalibration impact these results? How do deviations from the Gaussian noise model impact the learning curves of the algorithms? How accurately can the system parameters be calibrated, and how does that impact precision? Is the (approximated) radial basis kernel the most appropriate covariance function? How should increased sample density be traded off with sampling in under-sampled locations, give limited resources to transmit samples?

Although we have also made some initial theoretical progress in automatically calibrating the bias of sensors in the sensing system using Gaussian process models [6], many questions also remain in this area. How does the automatic calibration hold up with a large, real data set. What is rate of drift of the calibration of the sensors? How much should we trade off calibrating vs. super-sampling? How can we infer the gain error of sensors?

Another significant obstacle to ubiquitous and personal sensing using mobile phones is obstruction of the airflow to the sensor (i.e. because the phone is in the user's pocket or purse). How can we detect this situation? Can it be compensated for, or do we need to discard samples taken in such a situation? In a related question, how can we detect indoor vs. outdoor environments. We have done some promising initial experiments using the microphone of the phone to classify the user's environment based on ambient noise, but these efforts need to be fleshed out.

Finally, many issues remain surround the end applications of the data. Can users be guided to safety in an emergency based on their position and the inferred position of a plume? How should data be visualized? How can it be anonymized while remaining sufficiently useful to various types of end users?

VII. Conclusion

Although many questions remain to be answered before we can build a working sensor system based on sensors integrated into mobile phones, we are encouraged by these results. We believe that mobile sensing has the potential to provide the platform for building the largest scientific instrument ever made: one with a dynamic range wide enough to construct an accurate image of the impact that humans have on their environment at a societal scale while also being able to examine an individual's exposure to a specific element at a specific place and time. Until now, no sensing system has been able to do this, and we believe that the potential benefits to society are enormous.

VIII. Acknowledgments

We would like thank the Intel CommonSense team for their help and support building prototype sensors. This work was partly funded by Intel Research and the National Science Foundation.

References

[1] AirNow, "AirNow: Quality of air means quality of life," http://www.airnow.gov/.

[2] R. Honicky, E. Brewer, E. Paulos, and R. White, "N-SMARTS: Networked Suite of Mobile Atmospheric Real-Time Sensors," in *Proceedings of Networked Systems for Developing Regions 2008*, 2008.

[3] M. R. Beychok, *Fundamentals of stack gass dispersion*. Milton R. Beychok, 2005.

[4] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. MIT Press, 2006.

[5] P. Sollich, "Learning curves for Gaussian processes," in *Proceedings of the 1998 conference on advances in neural information processing systems II*. Cambridge, MA, USA: MIT Press, 1999, pp. 344–350.

[6] R. Honicky, "Automatic calibration of sensor-phones using gaussian processes," EECS Department, UC Berkeley, Tech. Rep. UCB/EECS-2007-34, 2007.