UrbanSense08 Panel Discussion

Increasing the precision of mobile sensing systems through super sampling

Hi there. I'm RJ Honicky from UC Berkeley. I'm interested in understanding how we can interpret data from implicit, passive, human out-of-the-loop sensing systems. Today I'm going to follow the guidance of the reviewers, and focus on some of the questions I raise in my paper, rather emphasize a show and tell of what we did. Here's my position: I believe that a large percentage of the mobile phones can and should be built with pollution sensors in them, but that to get there from here, there are some serious challenges that we need to address.

The state of the pollution sensing today, more or less, is that relatively few, expensive, high precision sensors are strategically placed around cities. Researchers are also flying around in modified airplanes and driving instrumented vehicles to get snapshots of pollution at a finer granularity or larger scale.

Now, I don't want to belabor the point to this audience, but just briefly: With the ubiquity and economies of scale of mobile phones, we have a unique opportunity. Wireless sensor networks have failed to deliver on their promise of ubiquitous, low cost sensing in large part because of two factors. First, energy. Second, economy of scale. Mobile phones already have an answer for both of these problems. I don't think I need to elaborate on that any more.

Also, since the network and processor is already built in to phones, all we need to add is a little MEMS device, and a little bit of plumbing, for a very marginal cost (at least optimistically speaking). I believe there are a variety of market mechanisms to absorb this cost, which we can discuss in the panel if someone asks.

<SLIDE>

Taking a brief detour, this is the room where the generators for an ISP I visited in Guinea Bissau are housed. These are converted truck engines, and the one with the big exhaust pipe in the foreground, inside the room, was very poorly tuned. The wall on the left, which you can't see, is open to the elements, so the exhaust does have a route to escape, but I took a 300ppm reading at the *door* to this room, so presumably the concentration of CO *inside* was much higher.

The point of all this is that people are working in some *very* dangerous circumstances and might not even realize it. Once we discovered the problem with this generator, they stop using it until they could tune it, and they also installed a CO sensor to monitor their staff. The bottom line is that people, particularly in developing countries, but in the industrialized world as well are particularly vulnerable to dangerous atmospheric pollution, but we believe they will do something about it if they have the right information.

So I believe that the cost is marginal, and the benefit is significant.

Here's my position, in detail: we can and should build a high precision, fine granularity, societal scale *sensing instrument* using mobile phones, but we first need to address the following fundamental technical problems

<SLIDE>

1) Cheap sensors are imprecise and inaccurate
2) Sampling will be non-uniform in space and time
3) Sensors will become mis-calibrated, and people are unlikely to explicitly calibrate them
4) People's actions and environments will bias the readings from the sensors
5) Tracking people's location and environment is a threat to privacy

By societal scale, I mean big enough and common enough to impact a whole society. This means that people from all walks of life need to participate, and that people from all parts of our cities need to participate. So a viable solution *has to* be cheap and accessible.

In my paper, I go into detail about one way to address the imprecision of cheap sensors, by exploiting the fact that sampling will tend to occur in locations where people congregate. I'm going to go over that a little bit right now.

<SLIDE>

Here's a test chamber we've built in which we can carefully control the concentration of poison gasses such as CO and NOx, as well as the humidity of the air. This allows us to test and calibrate the precision of our sensors, and measure and characterize the noise in our sensing system.

<SLIDE>

Looking at the signal from the sensors we're using, we see a pretty noisy signal. Here's a FFT of the raw signal from the sensor. There are a lot of spikes at the harmonics of 60hz. The sensor is picking these up over the air, not because of direct coupling to the AC power system.

<SLIDE>

Here's a histogram of the signal. If we clean it up with a low pass filter, then we get a nice Gaussian curve!

<SLIDE>

In this plot we can see the signal from one sensor (the lighter dots) and the *average* from six sensors (the darker dots). As you might expect, the variance of the average is significantly lower.

<SLIDE>

In fact, here's a plot of the variance versus the number of sensors. We can see that it follows the one over n curve that a Gaussian noise model predicts.

This is good news for us, because this means that we can increase the precision of the system significantly when sensors are close to one another by averaging their signals.

<SLIDE>

In fact, using a statistical method called a Gaussian process, we can take advantage of information from nearby sensors, proportional to their closeness to one another. This plot shows the variance of a signal in a simulation as the density of sensors "nearby" increases.

<SLIDE>

So, that's one way that we can take advantage of mobility to solve a problem with data we will get from mobile phones.

I also have some ideas about the other important problems that mentioned earlier. Automatic calibration will be crucial because sensors become miscalibrated over time, and people probably won't service them. So I've developed some simple techniques that can be applied to Gaussian processes that will automatically calibrate sensors when they're near one another by comparing their readings.

We have also been investigating how to interpolate values based on a physical model that includes wind and diffusion. Without modeling wind, our predictions about sparsely sampled areas will be poor, since diffusion is often slower than the wind speed.

Since people put phones in their pockets, drive cars and work indoors, we have to be able to detect these activities when we're taking samples so that we know how to interpret the data we gather. In some cases, we might have to throw the data away. In other cases, we might be able to label the data, and use them for tracking people's exposure to pollution verses their environment. We might even be able to recover the pollution levels in the environment if we understand the way, say, a person's pocket filters the pollution that the sensor is exposed to.

Without detecting these conditions, the noise in the signal will swamp the information, and we'll be stuck mostly with readings that might be useful for tracking personal exposure, but tell us very little about the environment.

Finally, Privacy is an area where I'm not very well versed, so I hope some of you all have some ideas about this.

<SLIDE>

You can find more information about my research on the NSMARTS web page. Thank you very much!