Toward a Bayesian Approach for Self-Tracking Personal Pollution Exposures

Debaleena Chattopadhyay
Department of Computer Science
University of Illinois, Chicago
debchatt@uic.edu

Sensors & Behaviuor Workshop '18
UbiComp/ISWC '18, October 8-12, 2018, Singapore

Abstract

Pollution exposure assessment at the population level is an established enterprise for environmental scientists and public health officials—but efforts to help individuals monitor and track their personal pollution exposures have just begun to garner research interest. Self-tracking pollution exposure is challenging for several reasons, including current limitations in sensor size, accuracy, and cost, frequent calibration requirements, and that people's daily activities often interfere with data quality in wearable sensing. The goal of this research is to develop a human-centered computing framework for the emerging field of personal pollution exposure assessment. To that aim, in this position paper, we propose a Bayesian approach to combine environmental sensing data from different spatiotemporal resolutions, such as from citywide national monitoring stations, neighborhood-wide lightweight sensing nodes, and personal wearables.

Author Keywords

Environmental sensing; pollution exposure profiling; self-tracking; pollution monitoring.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction

Ubiquitous computing has fundamentally changed the way we track our own behaviors. Recent breakthroughs in sensing technology, device miniaturization, and data analysis approaches are giving people access to data about their own behaviors, health indicators, thoughts, and feelings, that are orders of magnitude richer than any previously collected. One such health indicator is personal pollution exposure.

Monitoring pollution exposure is increasingly becoming important because of the many recognized health risks to the general population. Such efforts are commonplace in epidemiology [13], where measurements from central-site ambient monitors are used to model and estimate people's exposures to pollution [9]. Although high-end, a few fixed stations distributed city wide only offer a low spatiotemporal resolution, thereby, failing to account for several factors influencing personal exposures, such as higher exposures near localized emissions sources, disparities in how different outdoor pollutants infiltrate to indoor environments, human activity and mobility patterns, and indoor air quality (IAQ).

Recently, off-the-shelf IAQ monitoring technologies have gained prevalence among consumers; they provide household-level pollutant data, such as volatile organic compounds (VOC) and $PM_{2.5}$ [3]. While IAQ monitors are portable, they are designed to be stationary (e.g., placed on a living room side table [4]) and thus, limited toward accounting for human activity and mobility patterns. Hence, the upsurge of personal air pollution monitoring sensors that are carriable, handheld, or wearable. Although low-cost and suited for generating high-resolution pollution maps through participatory use, these sensors suffer from low selectivity. How to combat such measurement limitations of suboptimal pollution monitoring sensors is currently gaining research momentum [7]. Equally important but less explored is how to help individuals self-track their pollution exposures—beyond showing data from a few central-site ambient monitors [1].

For instance, several epidemiological studies have lately shown that long-term exposures to air and noise pollution are positively associated with mild-cognitive impairment (MCI) in older adults. MCI describes the stages between normal cognitive changes due to aging and early dementia, e.g., amnestic type of MCI reflects the onset of Alzheimer disease, and nonamnestic MCI has been linked to prodromal stages of vascular and other forms of dementia. Ubiquitous computing offers the promise of estimating pollution exposures: How much dust (particulate matter) did I inhale last month? When was I most exposed to ambient noise today? Am I currently exposed to a high level of air pollution? Should I take another route for my evening walk?

Background and Significance

Unfortunately, in the domain of self-tracking, our ability to help people monitor their personal pollution exposures lags substantially behind other health indicators, such as exercise, weight, food, sleep, or mood. Although pollution exposure assessment at the population level is an established enterprise for environmental scientists, personally tracking pollution exposures is challenging—especially by the vulnerable populations—for three main reasons. First, wearable, portable, environmental sensors are low-cost but suboptimal and sensitive to interference due to people's daily life activities, such as human skin emissions or textile emissions. Second, measurements from central-site monitors, the gold standard metric of exposure, are accurate but lack spatial and temporal resolution (e.g., 1-in-6-day schedule). Finally, current wearable pollution monitoring tools are essentially designed for citizen scientists and tech-savvy quantified-selfers. Thus, they may lack the design requirements to support at-risk communities, who need these tools the most.

Complementing the ongoing work on sensor calibration, we propose using sensor readings from different sensors toward inferring a personal exposure profile. For instance, national monitoring stations, fixed or mobile neighborhood-wide lightweight sensors [2, 6], and suboptimal wearable sensors [7]. These sensors provide data across a continuum of spatiotemporal resolution and measurement accuracy. To address the suboptimality of observed data from a wearable sensor, when constructing an inference about pollution exposures, we aim to use data from the other sources—combining different environmental sensing data using a generalized linear model (GLM) with random effects [10].

The choice of a GLM frees the requirement to use a normal distribution for modeling measurement errors and dependencies among response variables. From the literature, we can see that pollutants like $PM_{2.5}$ distribution tends to be positively skewed [11, 12]. As we are estimating the pollutant concentration at the same point in space and time, the observations from the three different monitoring sources are likely to be correlated. Thus, we will approach to solve the GLM as a problem in Bayesian inference. Contemporary Bayesian software makes tackling such robust linear regression computations straightforward. Next, we give an overview of the mathematical formulation for an example pollutant $PM_{2.5}$:

Let $x_1(s,t)$ be the logarithmic (log) pollutant value measured using EPA (US Environmental Protection Agency [15]) sites for any location s at time t. Similarly, $x_2(s,t)$ measured using AoT nodes (Chicago urban sensor network Array of Things [2]), and $x_3(s,t)$ measured using a personal wearable. We assume these observations are as followed:

$$x_1(s,t) = x(s,t) + \epsilon_1(s,t),\ x_2(s,t) = x(s,t) + \epsilon_2(s,t), \text{ and } x_3(s,t) = x(s,t) + \epsilon_3(s,t) \qquad \text{(Eq. 1)}$$

where $x(s,t)$ is the "true" unobserved value and ϵ_1, ϵ_2, and ϵ_3 the associated measurement errors. We then model

$$x(s,t) = \beta_0 + \beta_1 x_1(s,t) + \beta_2 x_2(s,t) + \beta_3 x_3(s,t) \qquad \text{(Eq. 2)}$$

To solve this as a Bayesian inference problem, we need to declare priors. Based on previous research [11], we can declare normal priors on log-$PM_{2.5}$ as:

$$x_1(s,t) \sim N(0,\sigma_1),\ x_2(s,t) \sim N(0,\sigma_2), \text{ and } x_3(s,t) \sim N(0,\sigma_3)$$

and assume uniform priors $Unif(0,5)$ for σ_1, σ_2, and σ_3 [12]. With these assumptions, if Φ_x is the collection of all parameters considered in the $PM_{2.5}$, we can finally compute the posterior predictive distribution as:

$$p(x(s_0,t_0) | x_1,x_2,x_3) \propto \int p(x(s_0,t_0) | x_1,x_2,x_3,\Phi_x)\ p(\Phi_x \mid x_1,x_2,x_3) d\Phi_x \qquad \text{(Eq. 3)}$$

The goal here is to determine what combinations of paraments in Φ_x are credible, given the monitoring data $D = \{x_1, x_2, x_3\}$, which is coming from the Bayes' rule:

$$p(\Phi_x \mid D) = \frac{p(D \mid \Phi_x)\, p(\Phi_x)}{\int p(D \mid \Phi_x) p(\Phi_x) d\Phi_x} \qquad \text{(Eq. 4)}$$

With that being our overarching approach toward self-tracking pollution exposures from multiple data sources, we now briefly discuss an ongoing project to demonstrate our research agenda's timely significance.

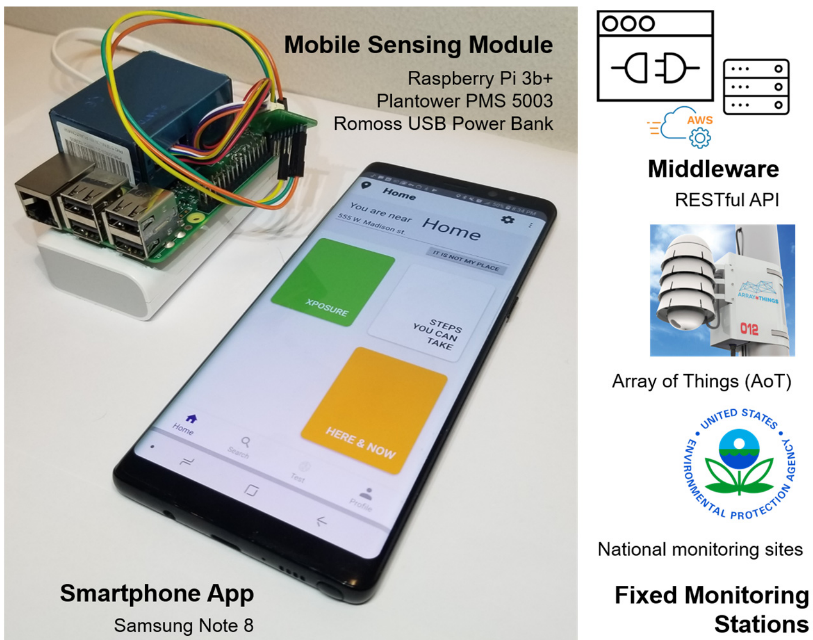

Figure 1. An overview of myCityMeter, a personal pollution exposure monitoring tool that collects data from three monitoring sources [4].

Example Project: myCityMeter

Epidemiological studies have shown that long-term exposures to air pollution, particularly $PM_{2.5}$, and ambient noise are positively associated with mild-cognitive impairment (MCI) in older adults [9, 14]. MCI de scribes the stages between normal cognitive changes due to aging and early dementia. To manage the environmental risk factors (ERFs) for MCI, older adults need to track their personal exposures to $PM_{2.5}$ and ambient noise.

We prototyped *myCityMeter*—a system that helps older adults track their daily and yearly exposures to $PM_{2.5}$ and noise, monitor current level of pollutants, take actions to combat pollutants' adverse effects, and test cognitive functions for an early MCI diagnosis [4]. Our system is composed of: (1) a mobile sensing module, (2) a middleware to access sensor readings from fixed measurement stations, and (3) a smartphone application (Figure 1).

Sensing module

The sensing module functions in conjunction with its companion smartphone (Figure 1). It is a wireless embedded sensing system with a computation and communication architecture based on the Raspberry Pi 3b+. It has a single board computer with a Bluetooth module, Wi-Fi, 1GB RAM, and 16GB storage. An off-the-shelf particle concentration sensor, Plantower PMS 5003, is mounted on the Pi, which uses a laser scattering principle to measure $PM_{2.5}$ per 5 s. The module is powered by a Romoss USB Power Bank (5V, 1A) and sensor data is transmitted to the Pi via SPI using the board's GPIO (Figure 1).

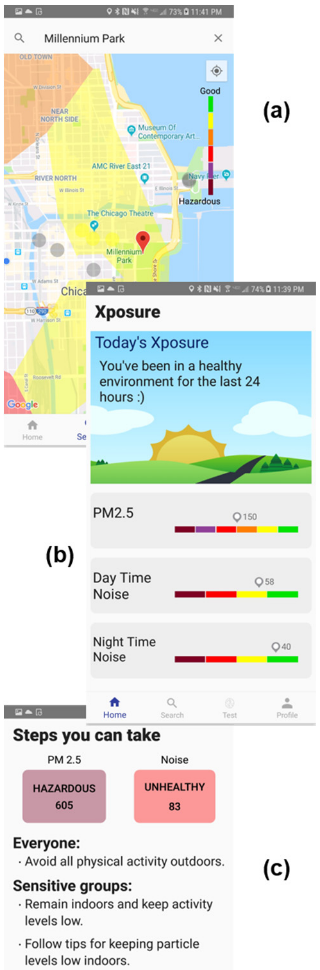

Figure 2. The phone application of myCityMeter, lookup (a), personal exposure profiles (b), and suggested actions (c).

The module is 3.2 x 2.2 x 1.7 inches and weighs about 100 gm. Pi sends the $PM_{2.5}$ readings directly to a cloud server when a Wi-Fi connection is available; or else readings are sent to the phone via a Bluetooth connection, and then the phone transmits the data to the server via LTE. If both Wi-Fi and Bluetooth are unavailable, readings are stored locally in the Pi. Each reading is geotagged (using the phone's GPS) and timestamped. The phone also records the ambient noise level and transforms the noise signal with A-weighting [dB(A)] before storing in the server. Our prototype currently works with an Android OS.

Middleware: Using other monitoring stations
Our middleware collects air quality data from EPA's four fixed stations in Chicago [14] and air and noise data from Chicago's AoT nodes [2]. AoT has 91 functional sensing nodes installed in different localities of Chicago, 16 providing $PM_{2.5}$ readings, and 63 ambient noise level.

The middleware of myCityMeter is hosted in an AWS cloud, which communicates with the phone client app using a RESTful API. Meanwhile, the client app in the phone records users' current location (if allowed) and sends it to the middleware. Currently, if a user is within 50 feet of an AoT node, a weighted average of the AoT readings—$G(x,y)$—is used along with the local readings to calculate the pollution exposure. The weights (w_i) are computed by first creating a Voronoi tessellation with n AoT nodes (x_i, y_i) and then computing a natural neighbor interpolation for the user's current location (x, y) as:

$$G(x,y) = \sum_{i=1}^{n} w_i f(x_i, y_i). \qquad \text{(Eq. 5)}$$

Companion smartphone application
The app allows users to look up pollution (Figure 2a) and their daily and yearly pollution exposures (Figure 2b). It suggests actions that users can take to avoid the adverse effects of pollutants (Figure 2c) and provides the Self-administered Gerocognitive Examination (SAGE; Figures 3a and 3b). Users can add their caregivers and provide different permissions (Figure 3c), such as to score and monitor their cognitive tests or access their current location.

Summary and Future Work
The novelty of this research proposal lies in the conceptualization and its accompanying computing methods for combining data from different environmental monitoring sources—at different spatiotemporal resolutions—to provide real-time personal pollution exposure assessment to individuals. To that end, we briefly described an early iteration of myCityMeter, a pollution exposure management tool for older adults.

As future work, we are pursuing three aims: (1) designing a wearable sensor module for a range of pollutants that leverages recent advancements in airborne particulate matter sensing, such as their inch-scale size; (2) implementing Bayesian learning for combining data from different environmental sensing sources; and (3) planning to conduct longitudinal studies to understand self-tracking pollution exposures by end-users, such as older adults and their caregivers.

Acknowledgments
UIC HCI lab students, Nina Sakhnini and Ja Eun Yu, developed the myCityMeter prototype.

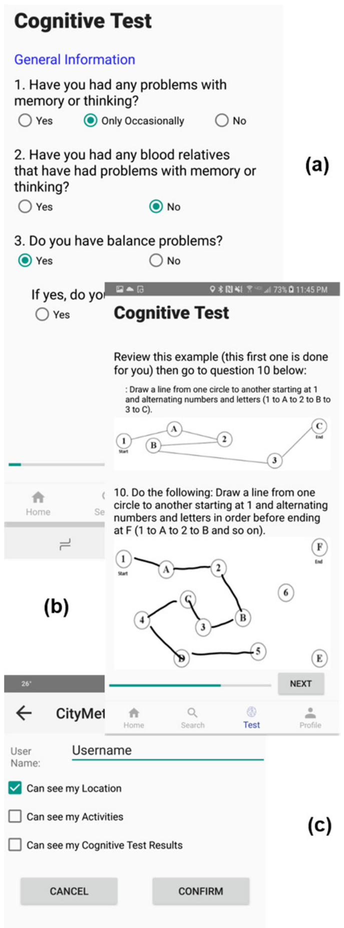

Figure 3. Screenshots of the SAGE cognitive test for diagnosing MCI (a, b) and giving caregivers monitoring permissions (c).

References

1. Ambiciti. Environmental awareness for all. Retrieved July 8, 2018 from http://ambiciti.io/
2. Array of Things. Retrieved July 8, 2018 from https://arrayofthings.github.io/
3. Biyi Fang, Qiumin Xu, et al. 2016. AirSense: an intelligent home-based sensing system for indoor air quality analytics. In Proc. UbiComp '16, 109–119. https://dx.doi.org/10.1145/2971648.2971720
4. Sakhnini, N., Yu, J., and Chattopadhyay, D. (2018). Poster: myCityMeter: Helping Older Adults Manage the Environmental Risk Factors for Cognitive Impairment. Ubicomp 2018 Adjunct. https://doi.org/10.1145/3267305.3267605
5. Sunyoung Kim and Eric Paulos. 2009. inAir: measuring and visualizing indoor air quality. In Proc. UbiComp '09, 81–84. https://dx.doi.org/10.1145/1620545.1620557
6. Yongsung Kim, Julien Eberle, Riikka Hanninen, Erol Can Un, et al. 2013. Mobile observatory: an exploratory study of mobile air quality monitoring application. In Proc. UbiComp '13 Adjunct, 733–736. https://dx.doi.org/10.1145/2494091.2495997
7. Balz Maag, Zimu Zhou, and Lothar Thiele. 2018. W-Air: Enabling Personal Air Pollution Monitoring on Wearables. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2, 1, Article 24, 25 pages. https://doi.org/10.1145/3191756
8. Douglas W. Scharre, Shu-Ing Chang, et al. 2010. Self-administered Gerocognitive Examination (SAGE): a brief cognitive assessment instrument for mild cognitive impairment (MCI) and early dementia. Alzheimer Disease & Associated Disorders, 24, 1: 64–71. https://doi.org/10.1097/WAD.0b013e3181b03277
9. Lilian Tzivian, Martha Dlugaj, Angela Winkler, et al. 2016. Long-term air pollution and traffic noise exposures and mild cognitive impairment in older adults: a cross-sectional analysis of the Heinz Nixdorf recall study. Environmental health perspectives, 124, 9:1361. https://dx.doi.org/10.1289%2Fehp.1509824
10. Zeger, S. L., & Karim, M. R. (1991). Generalized linear models with random effects; a Gibbs sampling approach. Journal of the American statistical association, 86(413), 79-86. https://doi.org/10.1080/01621459.1991.10475006
11. McMillan, N. J., Holland, D. M., Morara, M., & Feng, J. (2010). Combining numerical model output and particulate data using Bayesian space–time modeling. Environmetrics: The official journal of the International Environmetrics Society, 21(1), 48-65. https://doi.org/10.1002/env.984
12. Choi, J., Fuentes, M., & Reich, B. J. (2009). Spatial–temporal association between fine particulate matter and daily mortality. Computational statistics & data analysis, 53(8), 2989-3000. https://doi.org/10.1016/j.csda.2008.05.018
13. Jones, R., M., Jagai, J., Huang, Y-K., Chattopadhyay, D., Shi, X., and Wu, X. (2018). Data Infrastructure and Gateway for Environmental Health Research. In Proc. ISES-ISEE 2018 Joint Annual Meeting.
14. Power, M. C., Adar, S. D., Yanosky, J. D., & Weuve, J. (2016). Exposure to air pollution as a potential contributor to cognitive function, cognitive decline, brain imaging, and dementia: a systematic review of epidemiologic research. Neurotoxicology, 56, 235-253. https://doi.org/10.1016/j.neuro.2016.06.004
15. Illinois Environmental Protection Agency, Bureau of Air. (May 2018). Illinois Ambient Air Monitoring 2019 Network Plan. Retrieved from http://epa.illinois.gov/topics/air-quality/outdoor-air/air-monitoring/network/index on August 7, 2018