

# Meeting the data challenges of urban computing

September 14 2015, by Winnie Cui



Researchers proposed that they could estimate air quality at locations not covered by monitoring stations by analyzing the relationship between urban dynamics data.

Many people living in the People's Republic of China and Hong Kong have a new habit: they check the air pollution index before venturing outside. Air quality has deteriorated rapidly in China, with nitrogen dioxide and particulate matter levels frequently exceeding safety guidelines set by the World Health Organization.

While [poor air quality](#) clearly impacts public health, many cities have a dearth of air-quality monitoring stations: there are just 35 in Beijing and 15 in Hong Kong, for example. This lack of monitoring stations hinders evidence-based decision-making and leads to harsh criticisms of the transparency and public relevance of China's official [air pollution](#) index.

Since air pollution is highly location-dependent, a citywide air-quality

monitoring system would require building many additional monitoring stations, a solution that is prohibitively expensive. And so there is a grim reality: people may diligently check the [air pollution index](#), but they cannot really know the air quality in their specific locale.

"Do we have sufficient data to produce reliable air-quality metrics by using urban computing?" Professor Victor Li, chairman of information engineering and head of the department of Electrical and Electronic Engineering at the University of Hong Kong, posed this question to his research team. As luck would have it, while Li was challenging his researchers in Hong Kong, Microsoft Research Asia was launching a Call for Research Proposals dealing with urban computing. (Urban computing is a process of acquisition, integration, and analysis of big and heterogeneous data generated by a diversity of sources in urban spaces to address issues that major cities face, such as air pollution, excessive energy consumption, and traffic congestion.) Li promptly submitted a proposal, which came to the attention of Microsoft researcher Yu Zheng, who has conducted extensive research on urban computing and was also looking into the issues surrounding [urban air quality](#).

As Li and his researchers were detailing the correlations between air pollution and various urban dynamics—noting, for instance, that air quality and temperature are spatially correlated—Yu and his team had concluded that a variety of urban big data could be used compensate for the lack of monitoring stations. Li and Yu decided to collaborate and drive more in-depth research.

The joint team proposed that by analyzing the relationship between urban dynamics data, such as vehicular traffic and measured air quality, they could estimate air quality at locations not covered by monitoring stations. "We can infer real-time, fine-grained air quality information throughout a city based on the historical and real-time air quality data

reported by existing monitoring stations, combined with a variety of data sources we observe in the city," predicted Yu.

However, the researchers now grappled with the challenge of processing the massive volume of human dynamics data. A challenge initially imposed by a lack of data had become a problem of having too much data—a 180-degree swing from one extreme to the other!

But a solution to the problem of crunching the [big data](#) soon arrived, when Li's project received a Microsoft Azure for Research Award. As Julie Zhu, a doctoral student on the project team noted, "The program arrived at exactly the right time. We were just looking into building multi-node clusters."

Zhu and her colleagues attended Microsoft Azure training and quickly set up the new computing environment. Citing the enormous amounts of data provided by just one city, Shenzhen, Zhu described the value of Azure. "We need to collect and process about 1 terabyte per month of urban data on air quality, meteorological data, and traffic information, and so on," she said. "What's great about Microsoft Azure is that it goes way beyond the data storage. It integrates all the functionalities we need for data crawling, indexing, training, and visualization. Microsoft Azure truly enables us to do the real-time and scalable data processing."

Within months, the team had arrived at their initial results for Shenzhen. When Li and Yu presented their work at the 2014 Microsoft Research Asia Faculty Summit in Beijing, their novel approach generated great excitement, not only for how it processed the massive volumes of data, but also for finding the data sources in the first place. As Li commented, "All the data we used are from public channels. There are tons of data out there." Li and Yu are now working together to build a predictive model, which would address the initial dilemma of the paucity of [air-quality](#) monitoring stations.

Yu mused on how the challenge had morphed from insufficient data into a surplus of data. "We now know that the lack of data and big [data](#) do not conflict with each other. They co-exist in the same problem. What's important is to identify the dataset from various channels that are key contributors," Yu explained. To which Zhu, now a Microsoft Research intern, added, "You just need a good platform and tools to handle them."

Source: Microsoft

Citation: Meeting the data challenges of urban computing (2015, September 14) retrieved 7 August 2024 from <https://phys.org/news/2015-09-urban.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.