Inferring urban travel patterns from cellphone data

August 29 2016

Researchers from MIT and Ford Motor have developed a new computational system that uses cellphone location data to infer urban mobility patterns. Credit: Massachusetts Institute of Technology

In making decisions about infrastructure development and resource

allocation, city planners rely on models of how people move through their cities, on foot, in cars, and on public transportation. Those models are largely based on surveys of residents' travel habits.

But conducting surveys and analyzing their results is costly and time consuming: A city might go more than a decade between surveys. And even a broad survey will cover only a tiny fraction of a city's population.

In the latest issue of the *Proceedings of the National Academy of Sciences*, researchers from MIT and Ford Motor Company describe a new computational system that uses cellphone location data to infer urban mobility patterns. Applying the system to six weeks of data from residents of the Boston area, the researchers were able to quickly assemble the kind of model of urban mobility patterns that typically takes years to build.

The system holds the promise of not only more accurate and timely data about urban mobility but the ability to quickly determine whether particular attempts to address cities' transportation needs are working.

"In the U.S., every metropolitan area has an MPO, which is a metropolitan planning organization, and their main job is to use travel surveys to derive the travel demand model, which is their baseline for predicting and forecasting travel demand to build infrastructure," says Shan Jiang, a postdoc in the Human Mobility and Networks Lab in MIT's Department of Civil and Environmental Engineering and first author on the new paper. "So our method and model could be the next generation of tools for the planners to plan for the next generation of infrastructure."

To validate their new system, the researchers compared the model it generated to the model currently used by Boston's MPO. The two models accorded very well.

"The great advantage of our framework is that it learns mobility features from a large number of users, without having to ask them directly about their mobility choices," says Marta Gonzalez, an associate professor of civil and [environmental engineering](link) (CEE) at MIT and senior author on the paper. "Based on that, we create individual models to estimate complete daily trajectories of the vast majority of mobile-phone users. Likely, in time, we will see that this brings the comparative advantage of making urban transportation planning faster and smarter and even allows directly communicating recommendations to device users."

Joining Jiang and Gonzalez on the paper are Daniele Veneziano, a professor of CEE at MIT; Yingxiang Yang, a graduate student in CEE; Siddharth Gupta, a research assistant in the Human Mobility and Networks Lab, which Gonzalez leads; and Shounak Athavale, an information technology manager at Ford Motor's Palo Alto Research and Innovation Center.

Model building

The Boston MPO's practices are fairly typical of a major city's. Boston conducted one urban mobility survey in 1994 and another in 2010. Its current mobility model, however, still uses the data from 1994. That's because it's taken the intervening six years simply to sort through all the data collected in 2010. Only now has the work of organizing that data into a predictive model begun.

The 2010 survey asked each of 25,000 residents of the Boston area to keep a travel diary for a single day. From those diaries, combined with census data and information from traffic sensors, the MPO attempts to model the movements of 3.5 million residents of the greater Boston area.

While the MIT researchers had access to much more data—six weeks' worth from each of 1.92 million residents—it was less complete.

Cellphone records report only the locations at which users place calls or access the Internet. The researchers had to discard 25 percent of their data because it was too scanty.

From the rest, however, their algorithm was able to infer patterns of activity that recurred over the course of the six-week period. To piece together a picture of a cellphone user's day, the algorithm makes a few assumptions. One is that the location from which a user departs in the morning and to which she returns at night is her home. Another is that the location of the longest recurring stays during weekday daytime hours indicates the user's workplace.

Finally, the algorithm assumes that the lengths of most people's workdays accord with national averages. For instance, if a given user makes phone calls from work only between the hours of 12 p.m. and 2 p.m., the system does not interpret that as evidence of a two-hour workday—unless that interpretation is corroborated by other data, such as regular calls from home at 11:30 a.m. and 2:30 p.m. The estimates of workday length are probabilistic, however; the model doesn't assume that people arrive at work at exactly the same time every morning.

Any locations other than work and home are treated alike. From the available data, the system builds a probabilistic mobility model for each user, breaking every day of the week into 10-minute increments. For each increment, the model indicates the likeliness of a location change, possible destinations, and amount of time likely to be spent at each of those destinations. The system then generalizes those probabilities across communities, on the basis of census data, and deduces cumulative traffic flows from the resulting probability map.

  More information: The TimeGeo modeling framework for urban motility without travel surveys, www.pnas.org/cgi/doi/10.1073/pnas.1524261113