



Photo: Thinkstock

Ride-sharing could cut cabs' road time by 30 percent

A new analytic framework enables analysis of GPS data on 150 million cab rides in New York City.

Larry Hardesty | MIT News Office
September 1, 2014

Press Inquiries

Cellphone apps that find users car rides in real time are exploding in popularity: The car-service company Uber was recently valued at \$18 billion, and even as it faces legal wrangle a number of companies that provide similar services with licensed taxi cabs have sprung up

What if the taxi-service app on your cellphone had a button on it that let you indicate that you were willing to share a ride with another passenger? How drastically could cab-sharing reduce traffic, fares, and carbon dioxide emissions?

Authoritatively answering that question requires analyzing huge volumes of data, which has been computationally feasible with traditional methods. But in today's issue of the *Proceedings of the National Academies of Sciences*, researchers at MIT, Cornell University, and the Italian National Research Council's Institute for Informatics and Telematic present a new technique that enabled them to exhaustively analyze 150 million trip records collected from more than 13,000 New York City cabs over the course of a year.

Their conclusions: If passengers had been willing to tolerate no more than five minutes in delays per trip, almost 95 percent of the trips could have been shared. The optimal combination of trips would have reduced total travel time by 40 percent, with corresponding

reductions in operational costs and carbon dioxide emissions.

“Of course, nobody should ever be forced to share a vehicle,” says Carlo Ratti, professor of the practice in MIT’s Department of Urban Studies and Planning (DUSP) and one of the paper’s coauthors. “However, our research shows what would happen if people have sharing as an option. This is more than a theoretical exercise, with services such as Uber Pool bringing these ideas into practice.”

On the fly

Finding the optimal combination of trips does require foreknowledge of trips’ starting times: For instance, a 30-minute trip the length of Manhattan might be combined with a 10-minute trip beginning 15 minutes later. But that kind of advance planning is unlikely if the passengers are using cellphone apps. So the researchers also analyzed the data on the assumption that only trips starting within a minute of each other could be combined. Even then, they still found a 3 percent reduction in total travel time.

“We think that with the potential of a 30 percent reduction in operational costs, there is plenty of room for redistributing these benefits to customers, because we have to offer them lower fares; to drivers, because we have to incentivize them to belong to this system; to companies and of course, there is a benefit for the community,” says Paolo Santi, a visiting scientist in DUSP and first author on the paper.

In fact, Santi says, the results of his and his colleagues’ analysis were so striking that they asked Cornell mathematician Steven Strogatz to review their methodology. Strogatz is a co-author on the paper, as are Ratti and postdoc Stanislav Sobolevsky, both of MIT’s Senseable City Lab. Rounding out the author list are Michael Szell, who was a postdoc in the Senseable City lab when the work was done and is now at Northeastern University, and Giovanni Resta a researcher at Santi’s home institution, the Institute for Informatics and Telematics.

In analyzing taxi data for ride-sharing opportunities, “Typically, the approach that was taken was a variation of the so-called ‘traveling-salesman problem,’” Santi explains. “This is the basic algorithmic framework, and then there are extensions for sharing.”

The traveling-salesman problem asks whether, given a set of cities and the travel times between them, there is a route that would allow a traveling salesman to reach all of them with some time limit. Unfortunately, the traveling-salesman problem is also an example — indeed perhaps the most famous example — of an **NP-complete** problem, meaning that even for moderate-sized data sets, it can’t (as far as anyone knows) be solved in a reasonable amount of time.

Graphic results

So Santi and his colleagues took a different approach. First, they characterize every taxi trip according to four measurements: the time and GPS coordinates of both the pickup and the dropoff. Then, for each trip, their algorithm identifies the set of other trips that overlap with it — the ones that begin before it ends. Then it determines whether the trip they're examining can be combined with any of those other trips without exceeding the delay threshold. On average, any given trip is “shareable” with about 100 other trips.

Next, the algorithm represents the shareability of all 150 million trips in the database as a graph. A graph is a mathematical abstraction consisting of nodes — usually depicted as circles — and edges — usually depicted as lines between nodes. In this case, the nodes represent trips and the edges represent their shareability.

The graphical representation itself was the key to the researchers' analysis. With that in hand, well-known algorithms can efficiently find the optimal matchings to either maximize sharing or minimize travel time.

The researchers also conducted experiments to ensure that their matching algorithm would work in real time, if it ran on a server used to coordinate data from cellphones running a taxi-sharing app. They found that, even running on a single Linux box, it could find optimal matchings for about 100,000 trips in a tenth of a second, whereas the GPS data indicated that, on average, about 300 new taxi trips were initiated in New York every minute.

Finally, an online application designed by Szell, [HubCab](#), allows people to explore the taxi data themselves, using a map of New York as an interface.

David Mahfouda, the CEO of the car- and taxi-hailing company Bandwagon, whose business model is specifically built around ride sharing, says that his company hired analysts to examine the same data set that Santi and his colleagues did.

“We did analysis of rides from LaGuardia Airport and were able to build really detailed maps around where passengers were headed from that high-density departure point,” he says. But he adds, “we definitely simplified the problem in order to focus on a particular real-world problem that we thought we could solve. Making the entire data set available on a queryable basis does seem like a significantly larger lift.”

Mahfouda says that his company is founded on the assumption that “a very significant number” of taxi rides are shareable. “But it's extremely validating to have MIT corroborate that and it's been useful to our business to point to external validation,” he says.

He adds that, at the 2014 Consumer Electronics Show in Las Vegas, Bandwagon ran a demonstration version of its service for conference attendees. Over a four-day period, he says

ride sharing saved \$18,000 in fares and operational costs and more than 1,000 pounds in carbon emissions. But, he says, it also saved passengers time.

“Something that doesn’t get mentioned a lot in this space is the amount of time that gets saved through ride consolidation,” he says. “A lot of times people think that you have to wait longer order to find a shared ride. But particularly in congestion situations, you can ride faster if you’re willing to share vehicles.”