9 July 2007

Could algorithms ease frustrations for both airlines and passengers?

NARRATOR:
Welcome to the summer travel season. We're pretty sure you'll want to get to the airport early to check-in, just to make things easier.

But one thing you may not be able to control is whether you end up on your flight or not. That's because as airline revenue margins get tighter, airlines have to continue to overbook flights to make up for any of the "no-shows" with refundable tickets. It's a quandary for the airlines – business travelers, in particular, often demand refundable tickets, but if they don't show up, those empty seats can add up to billions of dollars in lost revenue for a major airline unless they overbook the seats to fill up the plane.

Yet if everyone does show up for a flight, bumping passengers can also be an expensive proposition – whether in vouchers for free flights or lost future revenue from annoyed customers.

It's a problem especially suited to mathematicians to solve. Dr. Rick Lawrence is a math scientist who manages the predictive modeling department at IBM's Watson Research Center.

RICK LAWRENCE:
"Airlines need to be able to predict quite accurately what the no-show rate will be on a specific flight. And the conventional approach to that problem is to build a historical, statistical based model, where they look at past instances of this flight and compute effectively what was the mean no-show behavior for this flight – perhaps a different no-show behavior if it was departing on a weekday as opposed to a weekend, and various other things – but the bottom line is, it was simply looking at aggregate historical behavior of similar flights. It does not... the conventional method does not take into account the composition of the passengers that are booked on that flight."

NARRATOR:
So in a test, Rick Lawrence and other IBM mathematicians worked with an actual airline to see if they could improve the predictions. To do that, they got more than a million anonymous passenger records over a 10-week period, with each record labeled as either a "show" or "no-show." And then they built a machine-learning model that started to make some connections for each of these nameless passengers, beyond just whether they'd bought a refundable or nonrefundable ticket.

RICK LAWRENCE:
"So we would not see the past history of this specific passenger; however we would know certain information about this passenger and this specific booking. For example, we would know what channel they booked this ticket through. Did they go through a travel agent? Did they go online? We even knew the specific identity of the online-booking channel. We found a significant amount of predictive information in that feature alone. We would also look at other information. For example did they order a special meal? Were they flying with other people? How many connection flights were in their itinerary?"

NARRATOR:
They used the data from the ten-week period to "train" their model to predict which passengers were likely to show and which wouldn't.

To add to the accuracy, Dr. Lawrence's model also took into account the conventional, historical flight-based predictions the airline already used. So by combining the conventional flight prediction model with the passenger-level prediction and summing that up as a cabin-level prediction, the mathematicians could then compare how accurately the new model would perform against the conventional model alone.

RICK LAWRENCE:
"It's the statistical cross-validation that gives us a metric of how much better we're doing than the conventional, historical method. And we found in our analysis that we could identify no-shows with approximately two times (twice) the accuracy, that we were seeing in the conventional model."

NARRATOR:
Which, if applied, could cut down on both the lost revenue from no-show passengers and the frustrations experienced by the ones who now show up only to get bumped. However, Rick Lawrence sees even greater potential in an industry that's moving things other than people around, where the no-show rates can be as high as forty percent sometimes.

RICK LAWRENCE:
"One of the interesting problems that we see now is, instead of asking whether an airline passenger is going to show, is to ask in the shipping industry, is a container going to show up for a vessel for which it's been booked. Interestingly enough, it's a similar dynamic, but currently there are fewer penalties and fewer disincentives for customers of shipping lines to have their shipment show up for that vessel. The additional wrinkle which is quite interesting is the fact that here the containers in a vessel for example, of course, have many different dimensions, so the extension here is the fact that not only does that container have associated with it a probability of no-show, just like the airline passenger problem, it has associated with it dimensions and other characteristics – does it need to be refrigerated, for example? -- that impose an additional optimization component that is somewhat different than the scenario that we think about in the airline industry."

NARRATOR:
Solving that problem some day could help you a lot – since we all pay for shipping for most of the goods we buy, one way or another. But as far as your summer vacation goes? At the moment we can only say that you'd better hurry, or else you might miss your flight.

For Ideas from IBM and the ibm.com homepage, I'm Derek Baker.

IBM Corporation
New Orchard Road, Armonk, New York 10504, USA