Researcher discusses a new study on correlations among medical problems

July 7 2015, by Peter Dizikes

What can big data tell us about the predictability of medical conditions? A new study by MIT researchers published in the journal Scientific Reports digs into this question by looking at anonymous data from over

500,000 patients. Among the findings is that for 43 percent of the patients, the data can help predict which of 1,719 subcategories of diseases are most likely to occur next. The co-authors of the paper are Carlo Ratti, director of MIT's Senseable City Laboratory, and two former computer science researchers at the lab, Dominik Dahlem (who is the lead author) and Diego Maniloff. The data originated with General Electric, which collaborated with Senseable City on a 2011 project on visually plotting health care data. MIT News spoke with Ratti about the new study.

Q. What is your central finding in the new study?

A. The results are quite interesting: This is one of the first analyses of large data you get from using electronic health records, and it just became available. This is a big amount of data we got from General Electric. What we tried to look at is, when you go to see the doctor, you've got a certain [medical] history, and you're perhaps looking at a [medical] problem. When you look at that problem, is there any predictive power in the history that comes before? We looked at that from a pure computer science point of view—and it turns out there is predictive power.

Q. In the paper, you state that "shuffling individual disease histories only marginally degrades the predictability bounds." That is, certain diseases correlate with each other largely apart of the order in which they occur, is that right?

A. You might want to reshuffle it [a patient's history] over time, to see how the predictability changes. And what we found was that you can predict even if you shuffle. Which in a certain sense tells you there are a

series of diseases that occur together. … They are not necessarily developing in a strict order, but it's about a cluster of things that come together.

At the level of the individual, this allows you to compare the medical history to other people, and give additional information to the doctor. Doctors can get additional input from this analysis of the medical history. Of course this is what doctors already do—they look at the past in order to understand what might be the problem. But it's a mathematical way that guides you, gives you more [than] than you might get by going through [one patient's medical history].

Q. Your lab has a focus on applying data to urban issues. So what was the genesis of this research project on health care?

A. Our focus is looking at how information is changing our knowledge of cities. And information from medical records is a very important type of information we can use. The question came about, can we actually look at these time sequences and try to understand—from just an information-theory point of view, can we actually predict—what comes next?

That is one of the things we have started doing with the data, looking at the data over space, and yes, we can see differences between different regions. And really you start understanding that interplay, about the individual, and quantifying the environment around ourselves … and that then becomes something that leaves a signature in medical records. In some sense, looking at medical records and the environment in certain regions becomes very important.

This story is republished courtesy of MIT News

(web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology