Using Twitter to Predict the Influence of Lifestyle on Health

University of Rochester

Researchers at the University of Rochester [1] showed last year how Twitter can be used to predict how likely it is for a Twitter user to become sick. They have now used Twitter to model how other factors – social status, exposure to pollution, interpersonal interaction and others – influence health.

"If you want to know, down to the individual level, how many people are sick in a population, you would have to survey the population, which is costly and time-consuming," said Adam Sadilek, postdoctoral researcher at the University of Rochester. "Twitter and the technology we have developed allow us to do this passively, quickly and inexpensively; we can listen in to what people are saying and mine this data to make predictions."

Sadilek also explained that many tweets are geo-tagged, which means they carry GPS information that shows exactly where the user was when he or she tweeted.

Collating all this information allows the researchers to map out, in space and in time, what people said in their tweets, but also where they were and when they were there. By following thousands of users as they tweet and go about their lives, reserchers also could estimate interactions between two users and between users and their environment.

In a paper to be presented on Feb. 8 at the International Conference on Web Searching and Data Mining in Rome, Italy, Sadilek will show how their new model accounts for many of the factors that affect health and how it can complement traditional studies in life sciences. Using tweets collected in New York City over a period of a month, they looked at factors like how often a person takes the subway, goes to the gym or a particular restaurant, proximity to a pollution source and their online social status. They looked at 70 factors in total. They then looked at whether these had a positive, negative or neutral impact on the users' health.

Some of their results are perhaps not surprising; for example, pollution sources seem to have a negative effect on health. However, this is the first time this impact has been extracted from the online behavior of a large online population. The paper also reveals a broader pattern, where virtually any activity that involves human contact leads to significantly increased health risks. For example, even people who regularly go to the gym get sick marginally more often than less active individuals. However, people who merely talk about going to the gym, but actually never go (verified based on their GPS), get sick significantly more often. This shows that there are interesting confounding factors that can now be studied at scale.

The technology that Sadilek and his colleague Professor Henry Kautz have developed has led to a web application called GermTracker. The application color-codes users (from red to green) according to their health by mining information from their tweets for 10 cities worldwide. Using the GPS data encoded in the tweets the app can then place people on a map, which allows anyone using the application to see their distribution.

"This app can be used by people to make personal decisions about their health. For example, they might want to avoid a subway station if it's full of sick people," Sadilek suggested. "It could also be used in conjunction with other methods by governments or local authorities to try to understand outbursts of the flu."

It is now flu season and as the number of people with the flu across the U.S. increases, so do the number of people monitoring GermTracker. On some days in January 10,000 people visited the http://fount.in [2] website where the app is hosted.

The model that Sadilek and his colleagues developed is based on machine-learning. At the heart of their work is how they are training an algorithm to distinguish between tweets that suggest the person tweeting is sick and those that don't.

"It's like teaching a baby a new language," Sadilek said. He explained that they first generated a training set of data, 5,000 tweets that had been manually categorized and from which the algorithm can start to distinguish what words and phrases are associated with someone being sick. He added, "We need the algorithm to understand that someone who tweets 'I'm sick and have been in bed all day' should be characterized as sick, but 'I'm sick of driving around in this traffic' shouldn't be."

The application is also improving the algorithm. Every time someone goes onto the application and clicks on one of the colored dots that represent the tweeting users, they can see the specific tweet that led someone to be classified in a specific way. The application asks you to assess the tweet yourself and say whether you agree with the classification or not. This gets fed back into the algorithm, which continues to learn from its mistakes.

The authors have recently started two collaborations with researchers at the University of Rochester Medical Center. "In one effort, we are planning to link Twitter predictions to clinical influenza studies," said co-author Kautz, chair of the University's computer science department. "In another effort, we are working with faculty in the Department of Psychiatry and the School of Nursing on extending these techniques to monitor and measure factors impacting depression and other psychological disorders."

For more information visit http://www.rochester.edu [1]

Source URL (retrieved on *03/15/2014 - 7:51am*):
http://www.wirelessdesignmag.com/news/2013/02/using-twitter-predict-influence-

lifestyle-health?qt-most_popular=0&qt-blogs=0&qt-digital_editions=0

Links:
[1] http://www.rochester.edu
[2] http://fount.in