Tracking Twitter may enhance monitoring of food safety at restaurants

August 7 2013

A satellite image showing sample tweets that indicate food poisoning and marked as such by the nEmesis system developed by University of Rochester researchers. Credit: Adam Sadilek

A new system could tell you how likely it is for you to become ill if you visit a particular restaurant by 'listening' to the tweets from other restaurant patrons.

The University of Rochester researchers say their system, nEmesis, can help people make more informed decisions, and it also has the potential to complement traditional public health methods for monitoring food

safety, such as restaurant inspections. For example, it could enable what they call "adaptive inspections," inspections guided in part by the real-time information that nEmesis provides.

The system combines machine-learning and crowdsourcing techniques to analyze millions of tweets to find people reporting food poisoning symptoms following a restaurant visit. This volume of tweets would be impossible to analyze manually, the researchers note. Over a four-month period, the system collected 3.8 million tweets from more than 94,000 unique users in New York City, traced 23,000 restaurant visitors, and found 480 reports of likely food poisoning. They also found they correlate fairly well with public inspection data by the local health department, as the researchers describe in a paper to be presented at the Conference on Human Computation & Crowdsourcing in Palm Springs, Calif., in November.

The system ranks restaurants according to how likely it is for someone to become ill after visiting that restaurant.

"The Twitter reports are not an exact indicator – any individual case could well be due to factors unrelated to the restaurant meal – but in aggregate the numbers are revealing," said Henry Kautz, chair of the computer science department at the University of Rochester and co-author of the paper. In other words, a "seemingly random collection of online rants becomes an actionable alert," according to Kautz, which can help detect cases of foodborne illness in a timely manner.

nEmesis "listens" to relevant public tweets and detects restaurant visits by matching up where a person tweets from and the known locations of restaurants. People will often tweet from their phones or other mobile devices, which are GPS enabled. This means that tweets can be "geotagged": the tweet not only provides information in the 140 characters allowed, but also about where the user was at the time.

If a user tweets from a location that is determined to be a restaurant (by using the locations of 24,904 restaurants that had been visited by the Department of Health and Mental Hygiene in New York City), the system will continue to track this person's tweets for 72 hours, even when they're not geotagged, or when they are tweeted from a different device. If a user then tweets about feeling ill, the system captures the information that this person is now ill and had visited a specific restaurant.

The correlation between the Twitter data and the public inspection data means that about one third of the inspection scores could be reliably predicted from the Twitter data. The remainder of the scores show some disagreement. "This disagreement is interesting as the public inspection data is not perfect either," argued co-author Adam Sadilek, formerly a colleague of Kautz at Rochester and who is now at Google. "The adaptive inspections could reveal the real risk, which is currently hidden for both methods."

This work builds on earlier work by Kautz and Sadilek that used Twitter to find out how likely a specific user was to have flu-like symptoms, and also to find the influence of different lifestyle factors on health. At the heart of all this work is the algorithm that Sadilek developed to distinguish between tweets that suggest a person tweeting is sick and those that don't. This algorithm is based on machine-learning, or as Sadilek described it, "it's like teaching a baby a new language," only in this case it's a computational algorithm that is being taught.

In their new system, nEmesis, they brought in an extra layer of complexity to improve the algorithm; they used crowdsourcing. For any one person, it would be exhausting and time-consuming to look through thousands of tweets to categorize them. The end results might not even be very accurate if their judgment is not quite right.

Instead the researchers turned to Amazon's Mechanical Turk system to reach out to a crowd of readily available workers. These were paid small amounts of money to categorize some tweets that could then be used to train the algorithm. They ensured the pool of tweets they were going use was of high accuracy by having more than one worker look at each tweet and incentivizing the right answer by paying the workers when their answer agreed with that of the majority and deducting money when it didn't. The algorithm was then able to learn from the training samples how to spot tweets that show people that are likely to have foodborne illnesses.

Of course, the system only considers people who tweet, who might not even be a representative sample of the whole population or of the population visiting a restaurant. But the Twitter data can be used together with knowledge gained from other sources to detect foodborne illness in a timely manner. It provides an extra layer – a passive level of monitoring – which is cost-effective. And the information that nEmesis offers can benefit both Twitter and non-Twitter users.

Rochester researchers Sean Brennan, graduate student, and Vincent Silenzio, associate professor of psychiatry, are also both part of the team that worked on nEmesis.

Provided by University of Rochester