

Making sense of patterns in the Twitterverse

June 7 2013

If you think keeping up with what's happening via Twitter, Facebook and other social media is like drinking from a fire hose, multiply that by 7 billion – and you'll have a sense of what Court Corley wakes up to every morning.

Corley, a data scientist at the Department of Energy's Pacific Northwest National Laboratory, has created a powerful digital system capable of analyzing billions of tweets and other social [media messages](#) in just seconds, in an effort to discover patterns and make sense of all the information. His social media analysis tool, dubbed "SALSA" (SociAL Sensor Analytics), combined with extensive know-how – and a fair degree of chutzpah – allows someone like Corley to try to grasp it all.

"The world is equipped with human sensors – more than 7 billion and counting. It's by far the most extensive [sensor network](#) on the planet. What can we learn by paying attention?" Corley said.

Among the payoffs Corley envisions are emergency responders who receive crucial early information about [natural disasters](#) such as tornadoes; a tool that public health advocates can use to better protect people's health; and information about social unrest that could help nations protect their citizens. But finding those jewels amidst the effluent of digital minutia is a challenge.

"The task we all face is separating out the trivia, the useless information we all are blasted with every day, from the really good stuff that helps us live better lives. There's a lot of noise, but there's some very valuable

information too."

The work by Corley and colleagues Chase Dowling, Stuart Rose and Taylor McKenzie was named best paper given at the IEEE conference on Intelligence and Security Informatics in Seattle this week.

Immensely rich data set

One person's digital trash is another's digital treasure. For example, people known in social media circles as "Beliebers," named after entertainer Justin Bieber, covet inconsequential tidbits about Justin Bieber, while "non-Beliebers" send that data straight to the recycle bin.

The amount of data is mind-bending. In social media posted just in the single year ending Aug. 31, 2012, each hour on average witnessed:

- 30 million comments
- 25 million search queries
- 98,000 new tweets
- 3.8 million blog views
- 4.5 million event invites
- 7.1 million photos uploaded
- 5.5 million status updates
- The equivalent of 453 years of video watched

Several firms routinely sift posts on LinkedIn, Facebook, Twitter,

YouTube and other social media, then analyze the data to see what's trending. These efforts usually require a great deal of software and a lot of person-hours devoted specifically to using that application. It's what Corley terms a manual approach.

Corley is out to change that, by creating a systematic, science-based, and automated approach for understanding patterns around events found in social media.

It's not so simple as scanning tweets. Indeed, if Corley were to sit down and read each of the more than 20 billion entries in his data set from just a two-year period, it would take him more than 3,500 years if he spent just 5 seconds on each entry. If he hired 1 million helpers, it would take more than a day.

But it takes less than 10 seconds when he relies on PNNL's Institutional Computing resource, drawing on a computer cluster with more than 600 nodes named Olympus, which is among the Top 500 fastest supercomputers in the world.

"We are using the institutional computing horsepower of PNNL to analyze one of the richest data sets ever available to researchers," Corley said.

At the same time that his team is creating the computing resources to undertake the task, Corley is constructing a theory for how to analyze the data. He and his colleagues are determining baseline activity, culling the data to find routine patterns, and looking for patterns that indicate something out of the ordinary. Data might include how often a topic is the subject of social media, who is putting out the messages, and how often.

Corley notes additional challenges posed by social media. His programs

analyze data in more than 60 languages, for instance. And social media users have developed a lexicon of their own and often don't use traditional language. A post such as "aw my avalanna wristband @Avalanna @justinbieber rip angel pic.twitter.com/yldGVV7GHk" poses a challenge to people and computers alike.

Nevertheless, Corley's program is accurate much more often than not, catching the spirit of a social media comment accurately more than three out of every four instances, and accurately detecting patterns in social media more than 90 percent of the time.

Public health, emergency response

Much of the work so far has been around public health. According to media reports in China, the current H7N9 flu situation in China was highlighted on Sina Weibo, a China-based social media platform, weeks before it was recognized by government officials. And Corley's work with the social media working group of the International Society for Disease Surveillance focuses on the use of social media for effective public health interventions.

In collaboration with the Infectious Disease Society of America and Immunizations 4 Public Health, he has focused on the early identification of emerging immunization safety concerns.

"If you want to understand the concerns of parents about vaccines, you're never going to have the time to go out there and read hundreds of thousands, perhaps millions of tweets about those questions or concerns," Corley said. "By creating a system that can capture trends in just a few minutes, and observe shifts in opinion minute to minute, you can stay in front of the issue, for instance, by letting physicians in certain areas know how to customize the educational materials they provide to parents of young children."

Corley has looked closely at reaction to the vaccine that protects against HPV, which causes cervical cancer. The first vaccine was approved in 2006, when he was a graduate student, and his doctoral thesis focused on an analysis of [social media](#) messages connected to HPV. He found that creators of messages that named a specific drug company were less likely to be positive about the vaccine than others who did not mention any company by name.

Other potential applications include helping [emergency responders](#) react more efficiently to disasters like tornadoes, or identifying patterns that might indicate coming social unrest or even something as specific as a riot after a soccer game. More than a dozen college students or recent graduates are working with Corley to look at questions like these and others.

Provided by Pacific Northwest National Laboratory

Citation: Making sense of patterns in the Twitterverse (2013, June 7) retrieved 18 June 2024 from <https://phys.org/news/2013-06-patterns-twitterverse.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--