Carnegie Mellon study of Twitter sentiments yields results similar to public opinion polls

May 11 2010

Computer analysis of sentiments expressed in a billion Twitter messages during 2008-2009 yielded measures of consumer confidence and of presidential job approval similar to those of well-established public opinion polls, Carnegie Mellon University researchers report.

The findings suggest that analyzing the text found in streams of tweets could become a cheap, rapid means of gauging public opinion on at least some subjects, said Noah Smith, assistant professor of language technologies and machine learning in the School of Computer Science. But tools for extracting public opinion from social media text are still crude and social media remain in their infancy, he cautioned, so the extent to which these methods could replace or supplement traditional polling is still unknown.

"With seven million or more messages being tweeted each day, this data stream potentially allows us to take the temperature of the population very quickly," Smith said. "The results are noisy, as are the results of polls. Opinion pollsters have learned to compensate for these distortions, while we're still trying to identify and understand the noise in our data. Given that, I'm excited that we get any signal at all from social media that correlates with the polls."

The study findings will be presented May 25 at the Association for the Advancement of Artificial Intelligence's International Conference on Weblogs and Social Media in Washington, D.C.

In the study, Smith and his colleagues collected a billion microblog messages — averaging about 11 words each — posted to Twitter during 2008 and 2009. They used simple text analysis techniques to identify messages that pertained to the economy or to politics and then found words within the text that indicated if the writer expressed positive or negative sentiments.

Results regarding consumer confidence were compared with the Index of Consumer Sentiment (ICS) from Reuters/University of Michigan Surveys of Consumers and the Gallup Organization's Economic Confidence Index. Political sentiments regarding President Obama were compared with Gallup's daily tracking poll on presidential job approval and views regarding the 2008 U.S. presidential election were compared with a compilation of 46 different polls prepared by Pollster.com. The ICS, Gallup and Pollster.com measurements were all obtained from telephone surveys using traditional polling techniques.

The Twitter-derived sentiment measurements were much more volatile day-to-day than the polling data, but when the researchers "smoothed" the results by averaging them over a period of days, the results often correlated closely with the polling data, said Brendan O'Connor, a graduate student in Carnegie Mellon's Language Technologies Institute and first author of the study. Consumer confidence, for instance, followed the same general slide through 2008 and the same rebound in February/March of 2009 as was seen in the poll data. The researchers noted that the ICS and Gallup data had a correlation of 86 percent over the period; the Twitter-derived sentiments had between 72 percent and 79 percent correlation with the Gallup data, depending on the number of days averaged to smooth the data.

Likewise, both the Twitter-derived sentiments and the traditional polls reflected declining approval of President Obama's job performance during 2009, with a 72 percent correlation between them.

But the researchers found that their sentiment analysis did not correlate as well with election polling during 2008. For instance, increased mentions of "Obama" tended to correlate with rises in Barack Obama's polling numbers, but increased mentions of "McCain" also correlated with rises in Obama's popularity. Improved computational methods for understanding natural language, particularly the unusual lexicon of microblogs, will be necessary before Twitter feeds can be reliably mined to predict elections, the researchers concluded.

"The Web is so mainstream now that there's no question that the Web is representative somehow of the population," O'Connor said. But pinning down Web demographics is still difficult, he acknowledged, noting that Twitter traffic alone increased by a factor of 50 during the two-year span of the study.

Using computer programs to judge the sentiments of microblogs is fraught with potential error, but even with the crude tools used in this exploratory research, the accuracy is better than can be achieved by chance, O'Connor said. "The massive amount of data was crucial in making this work," he explained. "We don't need to get the sentiment of every individual right to understand sentiments in aggregate."

Improved natural language processing tools, as well as query-driven analysis and use of demographic and time stamp data available on some social media sites, could increase the sophistication and reliability of microblog analysis.

  More information: Download a copy of the paper here, www.cs.cmu.edu/~nasmith/

Provided by Carnegie Mellon University