Is your news article tweetworthy? Algorithm can predict that

August 3 2012, By Alison Hewitt

Roja Bandari.

(Phys.org) -- Reporters, bloggers and other media trying to boost their Twitter presence can learn a few tips from an unexpected discipline: electrical engineering.

Roja Bandari, a Ph.D. candidate in electrical engineering at UCLA, developed an algorithm that predicts with 84 percent accuracy whether a news article will be popular on Twitter or bomb on the social media site.

Bandari came up with the algorithm with two Hewlett Packard Lab researchers while she was working at HP as an intern. She shared the

paper in June at the 2012 International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media.

Using artificial intelligence methods, the trio examined millions of tweets linking to more than 40,000 news articles. More than four-fifths of the time, if the algorithm flags the article as "popular," more than 100 tweets will later link to it. A "medium tweet" article will still see a healthy 20-100 tweets, while unpopular articles can expect less than 20.

$$T^{0.45} = (0.2S - 0.1Ent_{ct} - 0.1Ent_{avg} + 0.2Ent_{max})^2$$

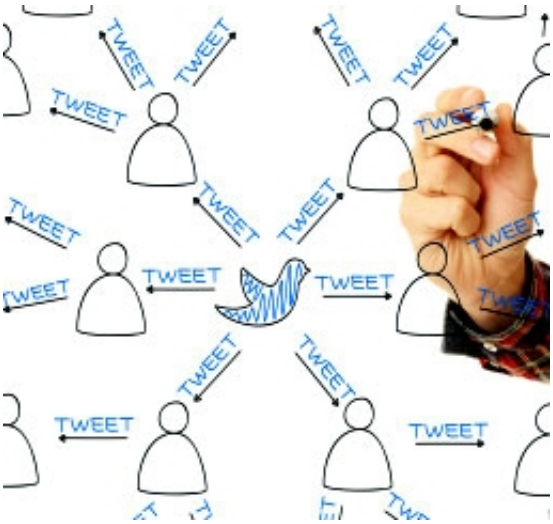
Trying to get your article to go viral on Twitter? This algorithm will solve all your problems — if you can understand it.

Bandari was already interested in examining how news items go viral from previous research she did at UCLA with Professor Vwani Roychowdhury, whose lab she still works in. In 2010, they examined how information propagated on Twitter during the 2009 election protests in Iran.

That inspired her to try and forecast the popularity of news articles on Twitter. Bandari looked at a handful of variables for each news article and found that the strongest predictor of an item's popularity was which news organization wrote the article. Certain media organizations seemed to draw tweets linking to their articles, whether from readers or the writers. For example, while the Christian Science Monitor's articles averaged 16 tweets — and some of its articles were never linked to at all on Twitter — the techie news site Mashable averaged 1,000 tweets per article, with its least-popular item still pulling in 360 tweets, the paper

noted.

“We found that the source of the article is the most important factor in whether an article will be popular on Twitter,” Bandari said. “And on Twitter, the sources that are successful are not what you might see elsewhere. Pretty much any article that Mashable shares on Twitter will get more tweets than an article from CNN. The Christian Science Monitor is a top source on Google, but it’s almost nonexistent on Twitter.”



Traditionally dominant media such as The New York Times, the Los Angeles Times and Reuters were all outshone on Twitter by news sources like TechCrunch, Mashable and the Huffington Post, she said.

The researchers also rated whether the articles had biased or emotional language by creating a “subjectivity classifier” — with a little help from Rush Limbaugh. The researchers used transcripts of the conservative

talk-show host as a model of highly subjective language. At the other end of the spectrum, they did the same with liberal commentator Keith Olbermann. CSPAN became the model for objective language.

To Bandari's surprise, the subjectivity of the language in the article was not a big predictor of popularity. Neither straightforward prose nor emotionally hot text had any apparent bearing on popularity, the researchers found.

"I had this very grim idea of news — that there are articles that get popular because they manipulate readers emotionally — and I was pleasantly surprised to see that doesn't play a role," she said. "So if you use heated language with the objective of increasing exposure, that's not necessarily going to work."

Topics were also an inconsistent measure of popularity, she said. Technology articles were the most popular on Twitter, but since articles can be classified under more than one subject, the researchers believe that makes accurate predictions difficult. And while the researchers expected that articles featuring famous people or well-known companies would be catnip on Twitter, the results were unclear, Bandari said.

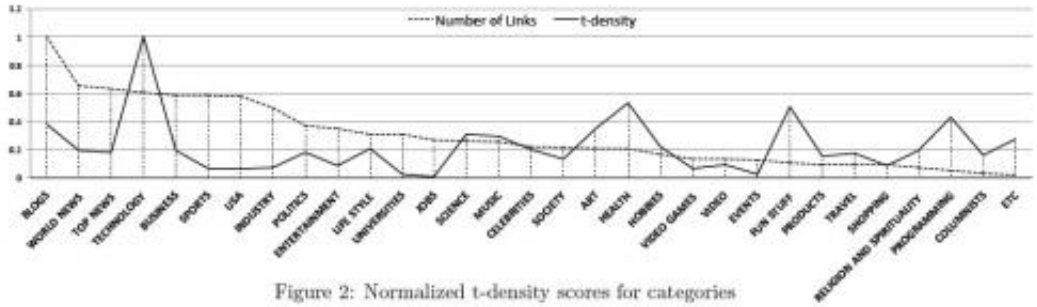


Figure 2: Normalized t-density scores for categories

The dotted line shows the amount of articles about each topic, and the solid line shows the amount of tweets per topic. Technology is clearly popular to tweet about. While health looks popular, Bandari noted that it and other topics on the

right side of the chart have a much smaller sample size of articles.

Nevertheless, she hopes the findings can provide some basic guidelines for journalists or bloggers interested in greater Twitter exposure.

To take advantage of the finding that the news brand matters, freelance writers might use this research to target Twitter-friendly outlets, Bandari speculated. “For example, if you’re a freelance reporter writing about tech, you don’t want to write for the Christian Science Monitor,” she said.

“If you were previously one of the dominant news sources, this shows that the rules have changed. You can’t rely on your old clout,” she said. “You need to compete with these new news outlets. It also shows that people are more interested in your content than in how emotional you make your argument. Hyped-up language doesn’t necessarily get you more exposure.”

Provided by University of California, Los Angeles