The Economist

Technology Quarterly: Q2 2011
In this Technology Quarterly

Monitor

Can Twitter predict the future?

Internet forecasting: Businesses are mining online messages to unearth consumers' moods—and even make market predictions

Jun 2nd 2011 | from the print edition

Belle Mellor

ONE day in 2008 an anonymous Twitter user posted a message: "I am certainly not bored. way busy! feel great!" That is all well and good, one might think, but utterly uninteresting to anyone besides the author and, perhaps, a few friends. Not so, according to Johan Bollen, of Indiana University Bloomington, who collected the tweet, along with plenty of others sent that day. All were rated for emotional content. Many proved similarly chirpy, scoring high on confidence, energy and happiness. Indeed, Dr Bollen reckons, on the day the tweet was posted, America's collective mood perked up a notch. When he and his team examined all the data for the autumn and winter of 2008, they found that Twitter users' collective mood swings coincided with national events. Happiness shot up around Thanksgiving, for example.

The idea of tapping web-based data to build a real-time measure of users' emotions and preferences is not new. Nor is that of using the results to predict their behaviour. Interest in internet forecasting was sparked by a paper published in 2009 by Hal Varian, Google's chief economist. He found that the peaks and troughs in the volume of Google searches for certain products, such as cars and holidays, preceded fluctuations in sales of those products. Other researchers have shown that searches for job-related terms are a good predictor of unemployment rates and that mentions of political candidates on Twitter correlate with electoral outcomes.

Dr Bollen spotted another curious correlation. When he compared trends in the national mood with movements of the Dow Jones Industrial Average (DJIA) he noticed that changes in one of the mood measure's seven components, anxiety, predicted swings in the share-price index. Spikes in anxiety levels were followed, around three days later, by dips in the price of shares. Why this happens remains unclear, but one possible explanation is that the falling prices were caused by traders' tendency to exit risky positions when feeling strung up.

Dr Bollen's algorithm, which he described in a paper published in February in the *Journal of Computational Science*, has been licensed to

Derwent Capital Markets, a hedge fund based in London. Derwent will use it to help guide the investments made with a £25m ($41m) fund that the firm hopes to launch in the next few months. Other funds are rumoured to be using similar tricks already. WiseWindow, a marketing firm based in Irvine, California, uses social-media activity to forecast demand for products. Its clients include Paramount Pictures and Belkin, a consumer-electronics firm.

All such initiatives face a problem, though. Humans excel at extracting meaning and sentiment from even the tiniest snippets of text, a task that stumps machines. To a computer, a tweet that reads "Feeling joyful after my trip to the dentist. Yeah, really" says that the author has been to the dentist and is now happy. Researchers have recently made strides in teaching machines to recognise such sarcasm, as well as double meanings or cultural references. In February Watson, a supercomputer devised by IBM, trounced two human champions at "Jeopardy!", an American quiz show renowned for the way its clues are laden with ambiguity, irony, riddles and puns. But, for the most part, processing natural language remains a challenge.

Then there is the question of how the new methods compare with existing approaches. Duncan Watts, an internet researcher who heads one of Yahoo!'s research labs in New York, likes to cite the example of weather forecasting in Santa Fe, New Mexico. A forecast that predicts sunshine every day would be correct 80% of the time, he quips. It would also be worthless, since locals already know that the town gets over 300 days of sun every year. In other words, to be useful a forecasting technique must add something to what is already known.

It is hard to know whether that is happening in finance, since hedge funds tend to keep mum about how successful (or not) particular forecasting algorithms are. So Dr Watts looked at areas, like films' box-office receipts, music and video-game sales, where some public data are available. How a film does at the box office, for instance, can be predicted based on bits of previously aggregated data: its production budget, the number of screens it opens on and wagers from HSX, a website which lets people place bets on movie revenues. Adding search data to the model, Dr Watts found, produced no improvement.

Search-related methods did, however, fare better when forecasting

video-game sales, where there are only two types of publicly available information to hand: ratings from reviews and, if the game is a sequel, sales figures for its predecessor. Forecasting models which added search data into the mix were much more accurate than those that did not. The same was true of music, another area where pre-packaged information is scarce.

Dr Watts thinks that simple search-volume forecasts will help spot consumer trends of this sort with increased precision. But the improvements they bring will be incremental. Sophisticated methods based on natural-language analysis of tweets, blogs, or Facebook pages, by contrast, hold greater disruptive potential. As users of social media grow accustomed to sharing highly personal information, apparently unfazed by market-research outfits like WiseWindow watching their every step, the feelings and intentions of hundreds of millions of people are there for data-hungry computers to see. And thanks to the likes of Dr Bollen, they are becoming increasingly legible.

from the print edition | Technology Quarterly