# Algos know more about us than we do about ourselves

John Dizard braves petabytes of internet babble for financial and commodities signals

**JOHN DIZARD**

To construct its Bank of Japan signal, Predata uses sources such as Twitter, the Japanese Wiki page for the BoJ and YouTube videos of BoJ governor Haruhiko Kuroda © EPA

NOVEMBER 24, 2017 John Dizard

When intelligence collectors and analysts take an interest in you, they usually start not by monitoring the content of your calls or messages, but by looking at the patterns of your communications. Who are you calling, how often and in what sequence? What topics do you comment on in social media?

This is called traffic analysis, and it can give a pretty good notion of what you and the people you know are thinking and what you are preparing to do. Traffic analysis started as a military intelligence methodology, and became systematic around the first world war. Without even knowing the content of encrypted messages, traffic analysts could map out an enemy "order of battle" or disposition of forces, and make inferences about commanders' intentions.

Traffic analysis techniques can also cut through the petabytes of redundant babble and chatter in the financial and political worlds. Even with state secrecy and the forests of non-disclosure agreements around "proprietary" investment or trading algorithms, crowds can be remarkably revealing in their open-source posts on social media.

Predata, a three-year-old New York and Washington-based predictive data analytics provider, has a Princeton-intensive crew of engineers and international affairs graduates working on early "signals" of market and political events. Predata trawls the open metadata for users of Twitter, Wikipedia, YouTube, Reddit and other social media, and analyses it to find indicators of future price moves or official actions.

I have been following their signals for a while and find them to be useful indicators. Predata started by creating political risk indicators, such as Iran-Saudi antagonism, Italian or Chilean labour unrest, or the relative enthusiasm for French political parties. Since the beginning of this year, they have been developing signals for financial and commodities markets.

# The 1-9-90 rule

1 per cent of internet users initiate discussions or content, 9 per cent transmit content or participate occasionally and 90 per cent are consumers or 'lurkers'

Jim Shinn, the chief executive of Predata and a former assistant secretary of defense and CIA officer, gives the example of the company's BoJ signal. For this, Predata collects the metadata from 300 sources, such as Twitter users, contested Wikipedia edits or YouTube items created by Japanese monetary policy geeks. Of those, at any time perhaps 100 are important, and 8 to 10 turn out to be predictive.

"If you can visualise a conversation about Japanese monetary policy, you can see it propagating across time following the 1-9-90 rule for internet conversations," Mr Shinn says. This holds that 1 per cent of users initiate discussions or content, 9 per cent transmit content or participate occasionally and 90 per cent are consumers or "lurkers". The actual percentages vary, but social media hierarchy takes the place of the formal chain of command or order of battle in traditional traffic analysis.

"The hierarchy can change every day," says Samuel Lee, a Predata engineer and financial analyst. "It is computationally very intensive to follow the leads and lags."

As Mr Shinn explains: "This is where you need some domain knowledge. It turns out that Twitter is pretty important for monetary policy, along with the Japanese-language Wiki page for the Bank of Japan, or, say, a YouTube video of [BoJ governor] Haruhiko Kuroda's cross-examination before a Diet parliamentary committee.

"Then you build a network of candidate discussions and look for the pattern those took before historical moves. The machine-learning algorithm goes back and picks the leads and lags between traffic and monetary policy events."

Typically, Predata's algos seem to be able to signal changes in policy or big price moves somewhere between 2 days and 2 weeks in advance. Unlike some academic Twitter scholars, Predata does not do systematic sentiment analysis of tweets or Wikipedia edits. "We only look for how many people there are in the conversation and comments, and how many people disagreed with each other. We call the latter the coefficient of contestation," Mr Shinn says.

The lead time for Twitter, Wiki or other social media signals varies from one market to another. Foreign exchange markets typically move within days, bond yields within a few days to a week,

and commodities prices within a week to two weeks. "If nothing happens within 30 days," says Mr Lee, "then we say we are wrong."

Since the beginning of the year, Predata's BoJ signal has been about 65 per cent accurate in calling major (two standard deviation) moves in the yen/dollar rate. According to Mr Lee, "Signals are more volatile on the FX side. If we say there will be a 2 STD move in the yen, then it has to happen within the 30 days or we don't give ourselves credit."

Certain politically related social media signals can lead commodity price moves. Predata follows the Arabic and Persian-language social media chatter about the Saudi-Iran rivalry intensely. Mr Shinn says: "We use relative volatility measures of the conversations. The drivers are the proxy conflicts between the two powers in Iraq, Lebanon and Yemen.

"It most obviously affects the oil market, say Brent crude 13 days after the [Riyadh-Tehran] signal is given. That has about a 64 per cent hit rate."

In the future, such signals can be tied to imagery, geographic location and spending behaviour. The algos know more about us than we do about ourselves.