Nowcasting the mood of the nation

Vast data-streams from social networks like Twitter and Facebook contain a people's opinions, fears and dreams. **Thomas Lansdall-Welfare**, **Vasileios Lampos** and **Nello Cristianini** exploit a whole new tool for social scientists.

Have you ever had the impression that everyone around you is stressed? Or is it just you? Answering a simple question like this can be very hard, as it would involve interviewing a large sample of people, and asking the right questions, in order to assess their levels of stress (unless you want to measure cortisol levels in faecal samples, as done in wild animal populations[1]).

The truth is that this is just one of many aspects of a population that are very difficult to measure or even detect. We often talk about macroeconomic quantities (such as the current level of inflation, or economic growth), forgetting that these quantities always refer to the past, since it takes several months to collect, aggregate and analyse the various economic indicators. Measuring the state of a society or an economy in real time is not an easy task. It is a task that certain practitioners call "nowcasting"[2].

My group has long been interested in nowcasting certain aspects of society by using the contents of social media. Active Twitter users in the UK number more than 10 million, Facebook users more than 31 million. Twitter, in particular and very helpfully, encourages its 200 million users worldwide to make their posts, commonly known as tweets, publicly available and tagged with the user's location. Tweets have another advantage: they tend to be of-the-moment, sent on impulse. They have immediacy; they reflect what the sender is feeling at the time, not what he or she feels looking back, a considered opinion from later.

All this generates a massive stream of status update and a new wave of experimentation and research. Can we find out something about Twitter and Facebook users – as a population – by analysing those posts?

Our first project back in 2010 had involved measuring the prevalence of flu-like symptoms in the general UK population, based on the contents of Twitter. We set up a system to read a large amount of tweets from the 54 largest cities in the UK, and a pipeline for the statistical analysis of text. We also collected the actual flu rates from the Health Protection Agency (HPA), so that we could use supervised computer-learning algorithms to map textual content to flu levels. We could have fed in key words and phrases, such as 'flu', or 'off work', or 'home in bed' and asked our text-recognition software to search for them. We did not have to: instead our machine learning algorithms worked out for themselves which words in the database of tweets occurred more often at times of elevated levels of flu. It generated a word cloud (Figure 1) where the size of the words is proportional to their weighting – to how strongly their presence indicates a flu epidemic – and where upside-down words have a negative weighting[3]. We ended up using a fancy version of linear regression to map word frequencies to flu levels, and we obtained very positive results: flu epidemics, it turns out, can be detected based on Twitter content. This was not entirely surprising, as it was known that flu levels can be "predicted" based on queries to Google[4]. (If you are feeling grotty, it is only natural to search the web for a description of your symptoms.) But it was definitely encouraging. Of course this is a case of nowcasting, not of forecasting.

In that study we tuned all the parameters of the algorithm automatically, to maximise the correlation of our indicator with the actual flu levels on the ground as published by the HPA. In this way, although we computed our indicator based on textual content in Twitter (therefore using a biased subset of UK population: city-dwelling Twitter users) we could make predictions about the general population. This would turn out to be impossible for our following study.

484 million tweets can reveal what a nation thinks of deaths of the famous, weddings of the famous – and of budget cuts

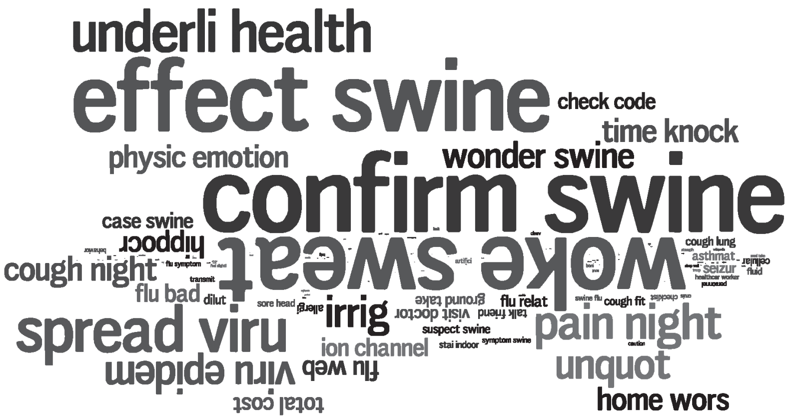

Figure 1. A word cloud automatically generated from Twitter traffic. The larger the word, the greater the correlation with flu epidemics. Upside-down words have negative correlations

We turned our attention to the issue of public mood, or sentiment. Our goal was to analyse the sentiment expressed in the collective discourse that constantly streams through Twitter. Or – as we called it – the mood of the nation.

We used tweets sampled from the 54 largest cities in the UK over a period of 30 months. There were more than 9 million different users, and 484 million tweets. It is important to notice that studies of this kind rely on very efficient methods of data management and text mining, which we have been refining for years, during our studies of news content[5], as well as social media content. Our infrastructure is based on a central database, and multiple independent modules that can annotate the data[6].

Notice also that the period we analysed goes from July 2009 to January 2012, a period marked by economic downturn and some social tensions. This will become relevant when analysing our findings.

There are standard methods in text analysis to detect sentiment: they are used mostly in marketing research, when analysts want to know the opinion of users of a certain camera, or viewers of a certain TV show. Each of the basic emotions (fear, joy, anger, sadness) is associated with a list of words, generated by a combination of manual and automatic methods, and successively benchmarked on a test set. This is called citation-sentiment analysis. We did not want to develop a new method for sentiment analysis, so we directly applied a standard one to the textual stream generated by UK Twitter users. We sampled the tweet-stream every 3 to 5 minutes, specifying location to within 10 km of an urban centre. Our word-list contained 146 anger words, 92 fear words, 224 joy words and 115 sadness words. They can be found at the WordNet-Affect website (`http://wordnet.princeton.edu`)[7].

In the flu project we had a "ground truth", of independently-measured flu cases. This time around we did not, as no one seems to be constantly measuring sentiment in the general population. This means that the methods and the conclusions will be of a different nature. Whereas in the flu project the list of keywords (whose frequency is used to compute the flu score) is discovered by our algorithm, with the goal of maximising correlation with the ground truth, in the mood project we had to feed the key words in ourselves – we got them from citation-sentiment analysis as mentioned above – and we have no ground truth to compare the result with.

By applying these tools to a time series of about 3 years of Twitter content we found that each of the four key emotions changes over time, in a manner that is partly predictable (or at least interpretable). We were reassured to find there was a periodic peak of joy around Christmas (Figure 2) – surely due to greetings messages – and a periodic peak of fear around Halloween, again probably due to increased usage of certain keywords such as 'scary'. These were sanity checks, which showed us that word-counting methods can provide a reasonable approach to sentiment or mood analysis. How far Christmas greetings accurately represent real joy, as opposed to duty and wishful thinking, is of course another question. We do not expect that a high frequency of the word 'happy' necessarily signifies happier mood in the population. Our measures of mood are not perfect, but these effects could be filtered away by a more sophisticated tool designed to ignore conventional expressions such as 'Happy New Year'. It is, however, a remarkable observation that certain days have reliably similar values in different years. This suggests that we have reduced statistical errors to a very low level.

But what came out most strongly is the strong transition, towards a more negative mood, that started in the week of October 20th, 2010. This was the week that the Prime Minister Gordon Brown announced massive cuts in public spending. It was a clear change point that we could validate by a statistical test. It was, if you like, the moment that people realised that austerity was not just for others; it would be affecting their own lives too. The effects of that major shift in collective mood are still felt today.

We also found a sustained growth in anger (Figure 4) in the weeks leading up to the summer riots of August 2011, when parts of London and several other cities across England suffered widespread violence, looting and arson.

It is interesting that the growth in anger seems to have started before the riots themselves, but this does not mean that we could

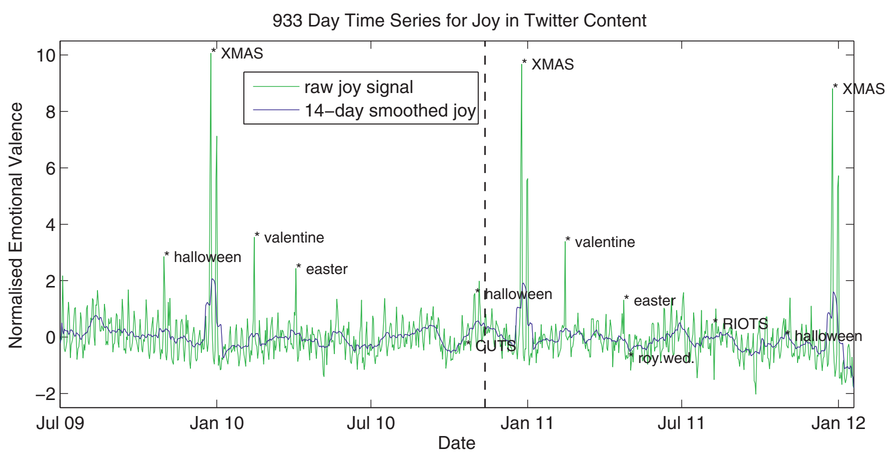

Figure 2. Plot of the time series representing levels of joy estimator over a period of 2½ years. Notice the peaks corresponding to Christmas and New Year, Valentine's day and the Royal Wedding

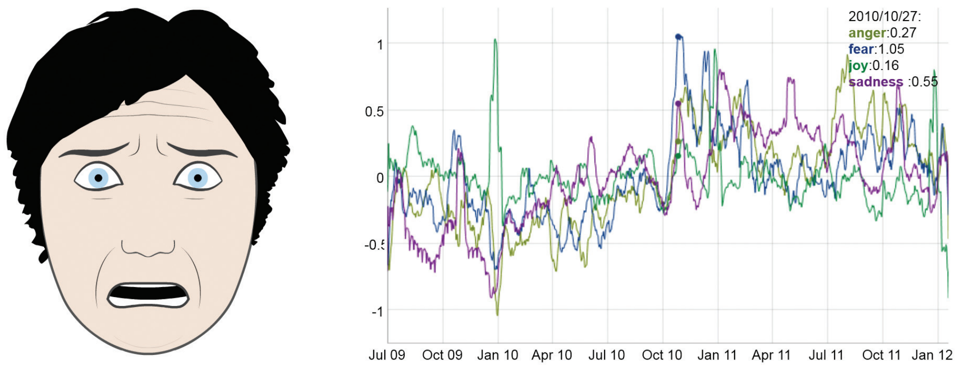

Figure 3. Visualisation of overall mood levels for the UK over 2½ years using timeline plots and the Grimace tool for facial expressions. The facial expression refers to October 27th, 2010. Visit `mediapatterns.enm.bris.ac.uk/mood`

have predicted them. Discovering an interesting correlation after the fact can be of great help to social scientists and other scholars, when interpreting those events, but is very different from predicting the events. There have been other increases in anger before, without this leading to any riots. As there is no official record of public mood, we need to be contented with finding correlations between trends in the time series of each emotion and events in the external world. We can find peaks of emotion for the death of Amy Winehouse, and of Osama Bin Laden; during the run-up to the Royal Wedding in April 2011 people felt calmer.

After the collection and the analysis part, we considered how to best visualise our results. With big data this is always a consideration. The data sets are so large, and the possible interactions they represent can be so complex, that graphic displays are becoming the norm. We are dealing with emotions; and we found an open source tool that represents emotions by a cartoon of a face whose expression depends on degrees of anger, joy, surprise, fear, sadness and disgust. It is called the grimace project (`http://grimace-project.net`), and we used it in conjunction with timelines. The end result can be used by the public as well as by researchers. Figure 3 is taken from our mood browser tool, which is live and interactive at `http://mediapatterns.enm.bris.ac.uk/mood/`. If you visit the site and drag the cursor along the time-line to October 2010, you will easily identify the week of the spending cuts: you will see the face suddenly wince.

There are some important considerations to make and lessons to learn, from the point of view of data analysis. The first is that the social sciences can now enter a data-driven phase, but this will require vast amounts of non-traditional data. The exploitation of big data will require the use of multiple tools, from different fields. Data management, data mining, text mining and data visualisation all seem to be as necessary as the statistical analysis part.

The second consideration is a caveat: since we did not choose the parameters of the mood system so as to correlate our score to the same score for the general UK population, we cannot claim that our mood scores were calibrated to compensate for the various and obvious biases we have in the data collection (unlike in the flu study). So all that we can claim – at best – is that we have measured the mood of city-dwelling Twitter users. They tend to be young; they tend to be savvy and techo-literate; they are definitely a biased sample of the UK population, although a large one, since we included posts by more than 9 million individual users.

Finally, there is the obvious caveat that goes with every statistical study: correlations – as we all know – are not causations. Even if there was an increase in anger and fear after the spending cuts were announced, how do we know that this was due to the announcement? Many other factors could have caused it. This is where data analysis must stop, and the interpretation of social scientists must begin. But at least we have collected and digested 484 million tweets for them, so that they can focus on the relevant questions. Big data can change the way social science is performed, but will not replace statistical common sense.

References

1. Weingrill, T., Gray, D.A., Barrett, L. and Henzic, S.P. (2004) Fecal cortisol levels in free-ranging female chacma baboons: relationship to dominance, reproductive state and environmental factors. *Hormones and Behavior*, **45**(4), 259–269.
2. Giannone, D., Reichlin, L. and Small, D. (2008) Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, **55**(4), 665–676.
3. Lampos, V. and Cristianini, N. (2011) Nowcasting events from the Social Web with statistical learning. *ACM Transactions on Intelligent Systems and Technology*, **3**(4).
4. Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S. and Brilliant, L. (2009) Detecting influenza epidemics using search engine query data. *Nature*, **457**(7232), 1012–1014.
5. `http://mediapatterns.enm.bris.ac.uk`
6. `http://www.tijldebie.net/system/files/SIGMOD_11_demo_Ilias.pdf`
7. Strapparava, C. and Valitutti, A. (2004) WordNet-Affect: an affective extension of WordNet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation* (LREC 2004), Lisbon, May, pp. 1083–1086.

Thomas Lansdall-Welfare, Vasileios Lampos and Nello Cristianini are at the Intelligent Systems Laboratory at the University of Bristol.

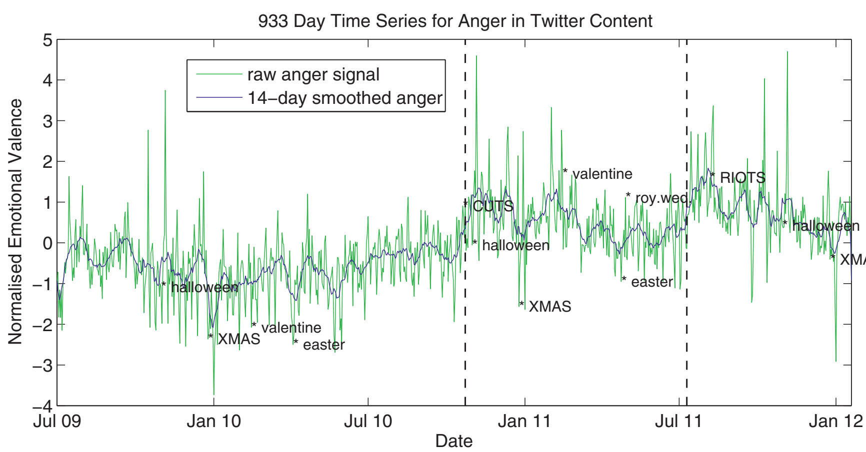

Figure 4. Plot of the time series for anger estimator over 2 and a half years. Notice visible change points corresponding to spending cuts and riots