

Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena

Johan Bollen

School of Informatics and Computing
Indiana University

Huina Mao

School of Informatics and Computing
Indiana University

Alberto Pepe

Center for Astrophysics
Harvard University

Abstract

We perform a sentiment analysis of all tweets published on the microblogging platform Twitter in the second half of 2008. We use a psychometric instrument to extract six mood states (tension, depression, anger, vigor, fatigue, confusion) from the aggregated Twitter content and compute a six-dimensional mood vector for each day in the timeline. We compare our results to a record of popular events gathered from media and sources. We find that events in the social, political, cultural and economic sphere do have a significant, immediate and highly specific effect on the various dimensions of public mood. We speculate that large scale analyses of mood can provide a solid platform to model collective emotive trends in terms of their predictive value with regards to existing social as well as economic indicators.

Introduction

Microblogging is an increasingly popular form of communication on the web. It allows users to broadcast brief text updates to the public or to a selected group of contacts. Microblog posts, commonly known as *tweets*, are extremely short in comparison to regular blog posts, being at most 140 characters in length. The launch of Twitter in October 2006 is responsible for the popularization of this simple, yet vastly popular form of communication on the web. Users of these online communities use microblogging to broadcast different types of information. A recent analysis of the Twitter network revealed a variegated mosaic of uses (Java et al. 2007), including a) daily chatter, e.g., posting what one is currently doing, b) conversations, i.e., directing tweets to specific users in their community of followers, c) information sharing, e.g., posting links to web pages, and d) news reporting, e.g., commentary on news and current affairs.

Despite the diversity of uses emerging from such a simple communication channel, it has been noted that tweets normally tend to fall in one of two different content camps: users that microblog about themselves and those that use microblogging primarily to share information (Mor Naaman 2010). In both cases, tweets can convey information about the mood state of their authors. In the former case, mood

expressions are evident by an explicit “sharing of subjectivity” (Crawford 2008), e.g. “I am feeling sad”. In other cases, even when a user is not specifically microblogging about their personal emotive status, the message can reflect their mood, e.g. “Colin Powell’s endorsement of Obama: amazing. :)”. As such, tweets may be regarded as microscopic instantiations of mood. It follows that the collection of all tweets published over a given time period can unveil changes in the state of public mood at a larger scale.

An increasing number of empirical analyses of sentiment and mood are based on textual collections of data generated on microblogging and social sites. Examples are mood surveys of communication on Myspace (Thelwall, Wilkinson, and Uppal 2009), and Twitter (Thelwall et al. 2010). Some of these analyses are focused on specific events, such as the study focused on microbloggers’ response to the death of Michael Jackson (Kim et al. 2009) or a political election in Germany (Tumasjan et al. 2010), while others analyze broader social and economic trends, such as the relationship between Twitter mood and both stock market fluctuations (Bollen, Mao, and Zeng 2010) and consumer confidence and political opinion (O’Connor et al. 2010). The results generated via the analysis of such collective mood aggregators are compelling and indicate that accurate public mood indicators can be extracted from online materials. Using publicly available online data to perform sentiment analyses significantly reduces the costs, efforts and time needed to administer large-scale public surveys and questionnaires. These data and results present great opportunities for psychologists and social scientists.

In this article, we explore how public mood patterns, as evidenced from a sentiment analysis of Twitter posts published between August 1 and December 20, 2008, relate to fluctuations in macroscopic social and economic indicators in the same time period. On the basis of a large corpus of public Twitter posts we look specifically at the interplay between a) macroscopic socio-cultural events, such as the outcome of a political election, and b) the public’s mood state measured by a well-established six-dimensional psychometric instrument. With our analysis we attempt to identify a quantifiable relationship between overall public mood and social, economic and other major events in the media and popular culture.

Method

Data and instrument

Our study is based on two data sources:

1. a timeline of important political, cultural, social, economic, and natural events occurred between August 1 and December 20, 2008.
2. a corpus of 9,664,952 tweets published by Twitter users in the same time period, and temporally distributed as shown in Figure 1.

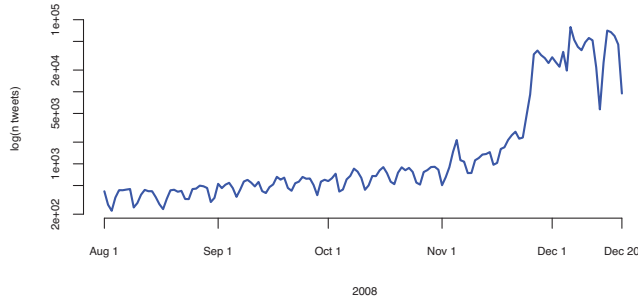


Figure 1: Volume of tweets: August 1 to December 20, 2008.

We develop and employ an extended version of a well established psychometric instrument, the Profile of Mood States (POMS) (McNair, Looor, and Droppleman 1971). POMS measures six individual dimensions of mood, namely *Tension*, *Depression*, *Anger*, *Vigour*, *Fatigue*, and *Confusion*. In its traditional usage, the POMS is not intended for large-scale textual analysis. Rather, it is a psychometric questionnaire composed of 65 base terms. Respondents to the questionnaire are asked to indicate on a five-point intensity scale how well each one of the 65 POMS adjectives describes their present mood state. The respondent's ratings for each mood adjective are then transformed by means of a scoring key to a 6-dimensional mood vector. The POMS is an easy-to-use, low-cost instrument whose factor-analytical structure has been repeatedly validated, recreated, and applied in hundreds of studies (McNair, Heuchert, and Shilony 2003). A number of reduced versions of the POMS have appeared in specialized literature, with the aim to condense the number of test terms and thus to reduce the time and effort on the part of human subjects to complete the POMS questionnaire (Cheung and Lam 2005). An extended version of the POMS, referred to as POMS-ex, has also been developed and validated in the literature (Pepe and Bollen 2008). This version of the instrument is not intended to be administered as a questionnaire to human subjects, but rather to be applicable to large textual corpora. POMS-ex extends the original set of 65 POMS mood adjectives to 793 terms, including synonyms and related word constructs, thus augmenting the possibility of matching terms in large data, such as online textual corpora.

Text processing and POMS scoring

Each individual tweet in our Twitter collection of 9,664,952 tweets is normalized and parsed before processing as fol-

lows:

1. Separation of individual terms on white-space boundaries
2. Removal of all non-alphanumeric characters from terms, e.g. commas, dashes, etc.
3. Conversion to lower-case of all remaining characters.
4. Removal of 214 standard stop words, including highly common verb-forms.
5. Porter stemming of all remaining terms in tweet.

This results in a subset of 1.1M normalized tweets which are POMS-scored. The POMS-scoring function $\mathcal{P}(t)$ maps each tweet to a six-dimensional mood vector $m \in \mathbb{R}^6$. The entries of m represent the following six dimensions of mood: *Tension*, *Depression*, *Anger*, *Vigour*, *Fatigue*, and *Confusion*.

The POMS-scoring function $\mathcal{P}(t)$ simply matches the terms extracted from each tweet to the set of POMS mood adjectives for each of POMS' 6 mood dimensions. Each tweet t is represented as the set w of n terms. The particular set of k POMS mood adjectives for dimension i is denoted as the set p_i . The POMS-scoring function, denoted \mathcal{P} , can thus be defined as follows:

$$\mathcal{P}(t) \rightarrow m \in \mathbb{R}^6 = [||w \cap p_1||, ||w \cap p_2||, \dots, ||w \cap p_6||]$$

The resulting mood vector m for tweet t is then **normalized** to produce the unit mood vector

$$\hat{m} = \frac{m}{||m||}$$

Time series production and normalization

We produce an aggregate mood vector m_d for the set of tweets submitted on a particular date d , denoted $T_d \subset T$ by simply averaging the mood vectors of the tweets submitted that day, i.e.

$$m_d = \frac{\sum_{t \in T_d} \hat{m}_t}{||T_d||}$$

The time series of aggregated, daily mood vectors m_d for a particular period of time $[i, i+k]$, denoted $\theta_{m_d}[i, k]$, is then defined as:

$$\theta_{m_d}[i, k] = [m_i, m_{i+1}, m_{i+2}, \dots, m_{i+k}]$$

A different number of tweets is submitted on any given day. Each entry of $\theta_{m_d}[i, k]$ is therefore derived from a different sample of $N_d = ||T_d||$ tweets. The probability that the terms extracted from the tweets submitted on any given day match the given number of POMS adjectives N_p thus varies considerably along the **binomial probability mass function**:

$$P(K = n) = \binom{N_p}{||W(T_d)||} p^{||W(T_d)||} (1-p)^{N_p - ||W(T_d)||}$$

where $P(K = n)$ represents the probability of achieving n number of POMS term matches, $||W(T_d)||$ represents the total number of terms extracted from the tweets submitted on day d vs. N_p the total number of POMS mood adjectives. Since the number of tweets per day varies considerably in

the time period under study, this leads to **systemic changes in the variance of $\theta_{m_d}[i, k]$ over time**, as shown in Fig. 2. In particular, the variance is larger when less tweets are submitted and lower as the number of tweets per day increases. **This effect makes it difficult to compare changes in the mood vectors of $\theta[i, k]$ over time.**

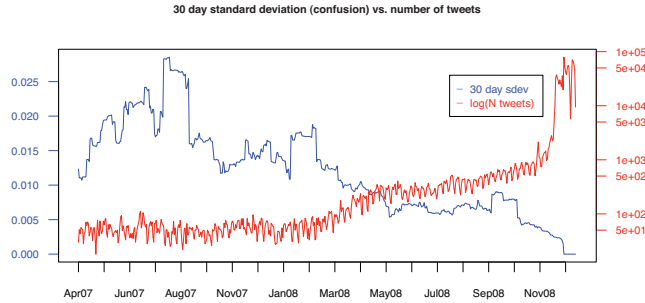


Figure 2: Standard deviation values of POMS Confusion scores within a 30 day window vs. the number of tweets submitted.

For this reason, we convert all mood values for a given day i to **z-scores** so that they would be normalized with respect to a local mean and standard deviation observed within the period $[i - k, i + k]$, i.e. **a sliding window of k days before and after the particular date**. The z-score of a mood vector \hat{m}_i for date i , denoted \tilde{m}_i , is then defined as:

$$\tilde{m}_i = \frac{\hat{m}_i - \bar{x}(\theta[i, \pm k])}{\sigma(\theta[i, \pm k])}$$

where $\bar{x}(\theta[i, \pm k])$ and $\sigma(\theta[i, \pm k])$ represent the mean and standard deviation of the time series within the local $[i, \pm k]$ days neighborhood of \hat{m}_i . When combined, the normalized mood vectors form the normalized time series:

$$\tilde{\theta}_{m_d}[i, k] = [\tilde{m}_i, \tilde{m}_{i+1}, \tilde{m}_{i+2}, \dots, \tilde{m}_{i+k}]$$

The effect of the z-score normalization is shown in Fig. 3 for the time series of the POMS confusion dimension over the course of 600 days¹. Where the time series is produced from a small number of tweets resulting in large swings in the un-normalized mood vectors, the magnitude of these swings is reduced by a commensurately high standard deviation. Where the time series is based on a larger sample of tweets, the standard deviation is smaller and thus a smaller swing in the un-normalized mood vectors is required to produce significant z-score fluctuations. As a result, the normalized time series fluctuates around a mean of zero and its fluctuations are expressed on a common scale, namely the standard deviation regardless of the number of tweets submitted on a particular date. This allows us to interpret the magnitude of the time series' fluctuations in terms of a common scale.

The above described normalizations results in a 153 day, 6-dimensional time series that fluctuates around a mean of

¹The graph is generated over a wider time frame than the period August 1, 2008 to December 20, 2008 under investigation to better illustrate this effect

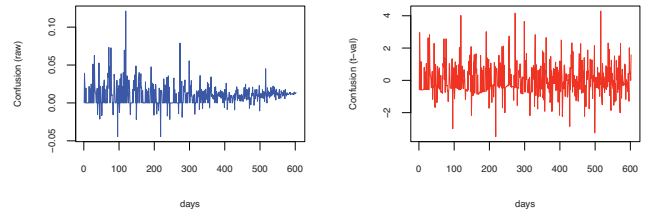


Figure 3: Raw POMS Confusion scores (left) vs. their z-scores normalization (right).

zero on a scale of 1 standard deviation representing fluctuations in the public mood from August 1 to December 20, 2008. Due to its normalization the time series highlights local, short-term deviations.

Results

Full results of this study can be found online as supplemental material, in the form of a chart². The chart displays, for the period under study, the timeline of socio-economic events and the time series extracted from our collection of tweets for each one of the POMS mood dimensions, z-score normalized. Shaded areas indicate the span of events that lasted for more than one day. Vertical lines originate in the time line's events and run across all mood dimensions to provide a visual frame of reference.

The period studied here, the latter half of year 2008, was marked by a number of remarkable socio-economic events of public interest: the U.S. presidential campaign and election, the failure of several large international banks, the Dow Jones dropping in value from above 11,000 points to less than 9,000, significant changes in the price of crude oil, and the official start of the deepest world-wide economic recession since World War II. The tumultuous nature of the timeline of socio-economic events in the period under study is reflected by the large fluctuations of the mood curves shown in the chart which exhibit large swings in value that range from several standard deviations below the mean to several standard deviations above the mean on a daily or weekly scale. By eyeballing the diagram, it is fairly easy to associate a number of major events in the timeline with some of the significant mood spikes. Notable examples are:

November 4 On U.S. election day, Tension skyrockets to over +2 standard deviations. The day after Vigour jumps from baseline levels to +3 standard deviations, while fatigue steadily drops to -2 standard deviations.

November 27 On U.S. Thanksgiving day, Vigour notably records a sharp increase from baseline levels to +4 standard deviations.

While these examples provide a useful yardstick to evaluate the efficiency of our sentiment tracking instrument against real world events, we explored macroscopic long-term effects of socio-economic indicators on general mood levels across longer periods of time. We calculated pairwise

²http://informatics.indiana.edu/jbollen/ICWSM11/publicmood_2008.pdf

Spearman Rank order correlations between each mood dimension by the day, thereby producing the 6×6 correlation matrix M , shown below,

$$M = \begin{bmatrix} & Ts & Cf & Vg & Ft & Ag & Dp \\ & 1.00 & 0.00 & 0.02 & -0.05 & 0.09 & 0.07 \\ & 0.00 & 1.00 & -0.04 & 0.00 & 0.06 & -0.02 \\ & 0.02 & -0.04 & 1.00 & -0.02 & 0.00 & -0.01 \\ & -0.05 & 0.00 & -0.02 & 1.00 & -0.06 & -0.01 \\ & 0.09 & 0.06 & 0.04 & -0.06 & 1.00 & 0.00 \\ & 0.07 & -0.02 & -0.01 & -0.01 & 0.00 & 1.00 \end{bmatrix}$$

where the abbreviations Ts, Cf, Vg, Ft, Ag, and Dp stand respectively for Tension, Confusion, Vigour, Fatigue, Anger, and Depression

Matrix M contains no statistically significant correlations for $N = 153$ which indicates that despite the tumult of events, the emotional response of the Twitter community was highly differentiated: none of the mood dimensions' values were statistically significantly correlated across all days in the period under investigation. In summary, although inconclusive with regards to the relation between long-term changes in socio-economic indicators, our results seem to suggest at least the following. First, events in the social, political, cultural and economical sphere do have a significant, immediate and highly specific effect on the various dimensions of public mood. These effects are short-lived as could be expected for mood states that are ephemeral and variable by definition. Second, economic events do seem to have an effect on public mood, but only to the degree that they correspond to rapid changes of economic indicators magnified by the media. Long-term changes seem to have a more gradual and cumulative effect. Third, continued negative drivers seem to have an effect on public mood but this effect may be manifested by short bursts of negative sentiment such as those observed on October 20, 2008. Finally, we would like to speculate that the social network of Twitter may highly affect the dynamics of public sentiment. Although we do not investigate the Twitter subscription network in this article, our results are suggestive of escalating bursts of mood activity, suggesting that sentiment spreads across network ties.

Conclusion

In this article, we perform a sentiment analysis of messages published on Twitter in the second half of 2008. We measure the sentiment of each tweet using an extended version of the Profile of Mood States (POMS) and compare our results to a timeline of notable events that took place in that time period. We find that social, political, cultural and economic events are correlated with significant, even if delayed fluctuations of public mood levels along a range of different mood dimensions. To conclude, we bring about the following methodological contribution: we argue that sentiment analysis of minute text corpora (such as tweets) is efficiently obtained via a syntactic, term-based approach that requires no training or machine learning. Sentiment analysis techniques rooted in machine learning yield accurate classification results when sufficiently large data is available for testing and training. However, minute texts such as microblogs may pose particular challenges for this approach.

Our method, which uses an analytic instrument rooted in decades of empirical psychometric research, proved a valid alternative to machine learning to detect public sentiment and associate its fluctuations with a timeline of socio-economic events.

References

- Bollen, J.; Mao, H.; and Zeng, X.-J. 2010. Twitter mood predicts the stock market. *Journal of Computational Science* 2(1):1–8.
- Cheung, S. Y., and Lam, E. T. C. 2005. An Innovative Shortened Bilingual Version of the Profile of Mood States (POMS-SBV). *School Psychology International* 26(1):121–128.
- Crawford, K. 2008. These foolish things: On intimacy and insignificance in mobile media. In Goggins, G., and Hjorth, L., eds., *Mobile Technologies: From Telecommunications to Media*. New York, NY, 10001: Routledge.
- Java, A.; Song, X.; Finin, T.; and Tseng, B. 2007. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web mining and Social Network Analysis*, 56–65. New York, NY, USA: ACM.
- Kim, E.; Gilbert, S.; Edwards, M.; and Graeff, E. 2009. Detecting Sadness in 140 Characters: Sentiment Analysis of Mourning Michael Jackson on Twitter. Technical report, Web Ecology Project, Boston, MA.
- McNair, D.; Heuchert, J. P.; and Shilony, E. 2003. *Profile of mood states. Bibliography 1964–2002*. Multi-Health Systems.
- McNair, D.; Lorr, M.; and Droppleman, L. 1971. Profile of Mood States.
- Mor Naaman, Jeffrey Boase, C.-H. L. 2010. Is it all About Me? User Content in Social Awareness Streams. In *Proceedings of the 2010 ACM conference on Computer Supported Cooperative Work*.
- O'Connor, B.; Balasubramanyan, R.; Routledge, B. R.; and Smith, N. A. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *Fourth International AAAI Conference on Weblogs and Social Media*.
- Pepe, A., and Bollen, J. 2008. Between conjecture and memento: shaping a collective emotional perception of the future. In *Proceedings of the AAAI Spring Symposium on Emotion, Personality, and Social Behavior*.
- Thelwall, M.; Buckley, K.; Paltoglou, G.; Cai, D.; and Kappas, A. 2010. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology* 61(12):2544–2558.
- Thelwall, M.; Wilkinson, D.; and Uppal, S. 2009. Data mining emotion in social network communication: Gender differences in MySpace. *Journal of the American Society for Information Science and Technology* In press.
- Tumasjan, A.; Sprenger, T. O.; Sandner, P. G.; and Welpe, I. M. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Fourth International AAAI Conference on Weblogs and Social Media*.