Why Are They Excited?
Identifying and Explaining Spikes in Blog Mood Levels

Krisztian Balog   **Gilad Mishne**   **Maarten de Rijke**
ISLA, University of Amsterdam
Kruislaan 403, 1098 SJ Amsterdam
`kbalog,gilad,mdr@science.uva.nl`

Abstract

We describe a method for discovering irregularities in temporal mood patterns appearing in a large corpus of blog posts, and labeling them with a natural language explanation. Simple techniques based on comparing corpus frequencies, coupled with large quantities of data, are shown to be effective for identifying the events underlying changes in global moods.

1 Introduction

Blogs, diary-like web pages containing highly opinionated personal commentary, are becoming increasingly popular. This new type of media offers a unique look into people's reactions and feelings towards current events, for a number of reasons. First, blogs are frequently updated, and like other forms of diaries are typically closely linked to ongoing events in the blogger's life. Second, the blog contents tend to be unmoderated and subjective, more so than mainstream media—expressing opinions, thoughts, and feeling. Finally, the large amount of blogs enables aggregation of thousands of opinions expressed every minute; this aggregation allows abstractions of the data, cleaning out noise and focusing on the main issues.

Many blog authoring environments allow bloggers to tag their entries with highly individual (and personal) features. Users of LiveJournal, one of the largest weblog communities, have the option of reporting their *mood* at the time of the post; users can either select a mood from a predefined list of common moods such as "amused" or "angry," or enter free-text. A large percentage of LiveJournal users tag their postings with a mood. This results in a stream of hundreds of weblog posts tagged with mood information per minute, from hundreds of thousands of users across the globe. The collection of such mood reports from many bloggers gives an aggregate mood of the blogosphere for each point in time: the popularity of different moods among bloggers at that time.

In previous work, we introduced a tool for tracking the aggregate mood of the blogosphere, and showed how it reflects global events (Mishne and de Rijke, 2006a). The tool's output includes graphs showing the popularity of moods in blog posts during a given interval; e.g., Figure 1 plots the mood level for "scared" during a 10 day period. While such graphs reflect some expected patterns (e.g., an increase in "scared" around Halloween in Figure 1), we have also witnessed spikes and drops for which no associated event was
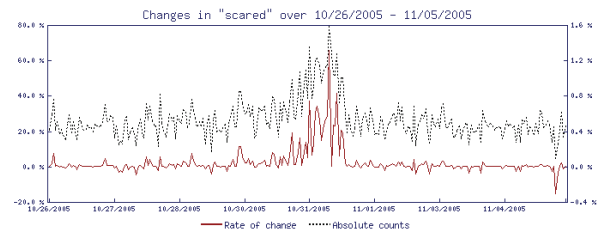

Figure 1: Blog posts labeled "scared" during the October 26–November 5, 2005 interval. The dotted (black) curve indicates the absolute number of posts labeled "scared," while the solid (red) curve shows the rate of change.

known to us. In this paper, we address this issue: we seek algorithms for identifying unusual changes in mood levels and explaining the underlying reasons for these changes. By "explanation" we mean a short snippet of text that describes the event that caused the unusual mood change.

To produce such explanations, we proceed as follows. If unusual spikes occur in the level of mood m, we examine the language used in blog posts labeled with m around and during the period in which the spike occurs. We interpret words

that are not expected given a long-term language model for m as *signals* for the spike in m's level. To operationalize the idea of "unexpected words" for a given mood, we use standard methods for corpus comparison; once identified, we use the "unexpected words" to consult a news corpus from which we retrieve a small text snippet that we then return as the desired explanation.

In Section 2 we briefly discuss related work. Then, we detail how we detect spikes in mood levels (in Section 3) and how we generate natural language explanations for such spikes (in Section 4). Experimental results are presented in Section 5, and in Section 6 we present our conclusions.

2 Related work

As to burstiness phenomena in web data, Kleinberg (2002) targets email and research papers, trying to identify sharp rises in word frequencies in document streams. Bursts can be found by searching periods when a given word tends to appear at unusually short intervals. Kumar et al. (2003) extend Kleinberg's algorithm to discover dense periods of "bursty" intra-community link creation in the blogspace, while Nanno et al. (2004) extend it to work on blogs. We use a simple comparison between long-term and short-term language models associated with a given mood to identify unusual word usage patterns.

Recent years have witnessed an increase in research on extracting subjective and other non-factual aspects of textual content; see (Shanahan et al., 2005) for an overview. Much work in this area focuses on recognizing and/or annotating evaluative textual expressions. In contrast, work that explores mood annotations is relatively scarce. Mishne (2005) reports on text mining experiments aimed at automatically tagging blog posts with moods. Mishne and de Rijke (2006a) lift this work to the aggregate level, and use natural language processing and machine learning to estimate aggregate mood levels from the text of blog entries.

3 Detecting spikes

Our first task is to identify spikes in moods reported in blog posts. Many of the moods reported by LiveJournal users display a cyclic behavior. There are some obvious moods with a daily cycle. For instance, people feel awake in the mornings and tired in the evening (Figure 2). Other moods show a weekly cycle. For instance, people drink more at the weekends (Figure 3).
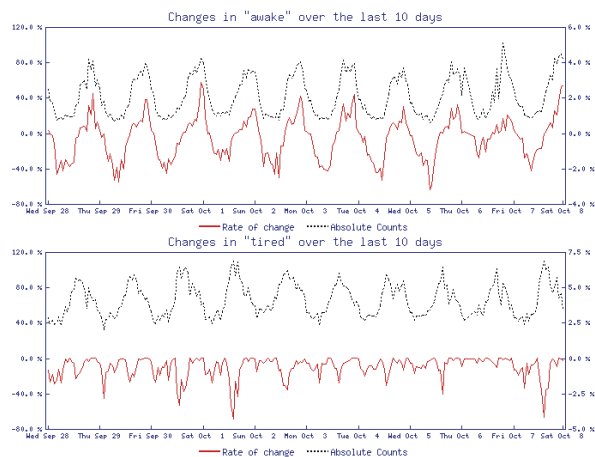

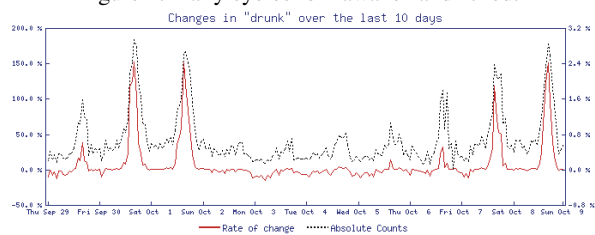Figure 2: Daily cycles for "awake" and "tired."

Figure 3: Weekend cycles for "drunk."

Our idea of detecting spikes tries to deal with these cyclic events and aims at finding global changes. Let $POSTS(mood, date, hour)$ be the number of posts labelled with a given mood and created within a one-hour interval at the specified date. Similarly, $ALLPOSTS(date, hour)$ is the number of all posts created within the interval specified by the date and hour. The ratio of posts labeled with a given mood to all posts could be expressed for all days of a week (Sunday, ..., Saturday) and for all one-hour intervals (0, ..., 23) using the formula:

$$R(mood, day, hour) = \frac{\sum_{DW(date)=day} POSTS(mood, date, hour)}{\sum_{DW(date)=day} ALLPOSTS(date, hour)},$$

where $day = 0, \ldots, 6$ and $DW(date)$ is a day-of-the-week function that returns $0, \ldots, 6$ depending on the date argument.

The level of a given mood is *changed* within a one-hour interval of a day, if the ratio of posts labelled with that mood to all posts, created within the interval, is significantly different from the ratio that has been observed on the same hour of the similar day of the week. Formally:

$$D(mood, date, hour) = \frac{\frac{POSTS(mood, date, hour)}{ALLPOSTS(date, hour)}}{R(mood, DW(date), hour)}.$$

If $|D|$ (the absolute value of D) exceeds a threshold we conclude that a spike has occurred, while

the sign of D makes it possible to distinguish between positive and negative spikes. The absolute value of D expresses the degree of the peak.

This method of identifying spikes allows us to look at a period of a few hours instead of only one, which is an effective smoothing method, especially if a sufficient number of posts cannot be observed for a given mood.

4 Explaining peaks

Our next task is to *explain* the peaks identified by the methods listed previously. We proceed in two steps. First, we discover features in the peaking interval which display a significantly different language usage from that found in the general language associated with the mood. Then we form queries using these "overused" words as well as the date(s) of the peaking interval and run these as queries against a news corpus.

4.1 Overused words

To discover the reasons underlying mood changes we use corpus-based techniques to identify changes in language usage. We compare two corpora: (1) the full set of blog posts, referred to as the *standard* corpus, and (2) a corpus associated with the peaking interval, referred to as the *sample* corpus.

To compare word frequencies across the two corpora we apply the log-likelihood statistical test (Dunning, 1993). Let O_i be the observed frequency of a term, N_i its total frequency, and $E_i = (N_i \cdot \sum_i O_i)/\sum_i N_i$ its expected frequency in corpus i (where i takes values 1 and 2 for the standard and sample corpus, respectively). Then, the log-likelihood value is calculated according to this formula: $-2 \ln \lambda = 2 \sum_i O_i \ln \left(\frac{O_i}{E_i} \right)$.

4.2 Finding explanations

Given the start and end dates of a peaking interval and a list of overused words from this period, a query is formed. This query is then submitted to (headlines of) a news corpus. A headline is retrieved if it contains at least one of the overused words and is dated within the peaking interval or the day before the beginning of the peak. The hits are ranked based on the number of overused terms contained in the headline.

5 Experiments

In this section we illustrate our methods with some examples and provide a preliminary analysis of their effectiveness.

5.1 The blog corpus

Our corpus consists of all public blogs published in LiveJournal during a 90 day period from July 5 to October 2, 2005, adding up to a total of 19 million blog posts. For each entry, the text of the post along with the date and time are indexed. Posts without an explicit mood indication (10M) are discarded. We applied standard preprocessing steps (stopword removal, stemming) to the text of blog posts.

5.2 The news corpus

The collection contains around 1000 news headlines that have been published in Wikinews (`http://www.wikinews.org`) during the period of July-September, 2005.

5.3 Case studies

We present three particular cases where an irregular behavior in a certain mood could be observed. We examine how accurately the overused terms describe the events that caused the spikes.

5.3.1 Harry Potter

In July, 2005, a peak in "excited" was discovered; see Figure 4, where the shaded (green) area indicates the "peak area."
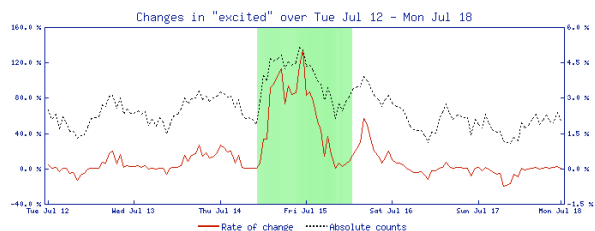

Figure 4: Peak in "excited" around July 16, 2005.

Step 1 of our peak explanation method (Section 4) reveals the following overused terms during the peak period: "potter," "book," "excit," "hbp," "read," "princ," "midnight." Step 2 of our peak explanation method (Section 4) exploits these words to retrieve the following headline from the news collection: "July 16. Harry Potter and the Half-Blood Prince released."

5.3.2 Hurricane Katrina

Our next example illustrates the need for careful thresholding when defining peaks (see Section 3). We show peaks in "worried" discovered around late August, with a 40% and 80% threshold. Clearly, far more peaks are identified with the lower threshold, while the peaks identified in the bottom plot (with the higher threshold), all appear to be clear peaks. The overused terms during the peak period include "orlean," "worri," "hurrican," "gas," "katrina" In
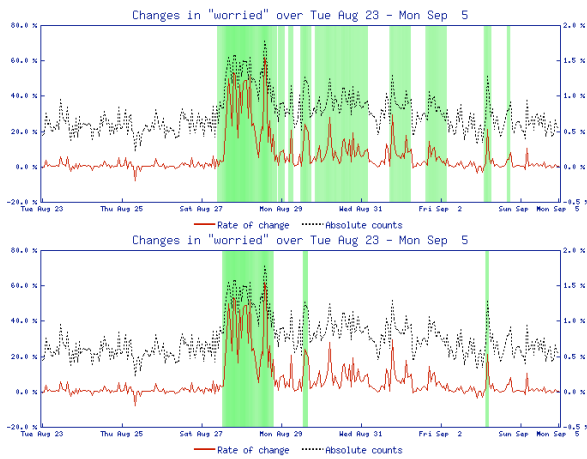
Figure 5: Peaks in "worried" around August 29, 2005. (Top: threshold 40% change; bottom: threshold 80% change)

Step 2 of our explanation method we retrieve the following news headlines (top 5 shown only):

(Sept 1) Hurricane Katrina: Resources regarding missing/located people

(Sept 2) Crime in New Orleans sharply increases after Hurricane Katrina

(Sept 1) Fats Domino missing in the wake of Hurricane Katrina

(Aug 30) At least 55 killed by Hurricane Katrina; serious flooding across affected region

(Aug 26) Hurricane Katrina strikes Florida, kills seven

5.3.3 London terror attacks On July 7 a sharp spike could be observed in the "sad" mood; see Figure 6; the tone of the shaded area shows the degree of the peak. Overused terms identified for this period include "london," "attack," "terrorist," "bomb,' "peopl", "explos." Consulting our news
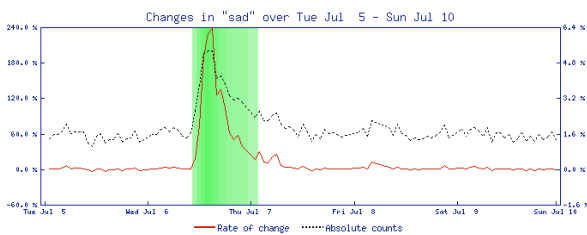

Figure 6: Peak in "sad" around July 7, 2005.

corpus produced the following top ranked results:

(July 7) Coordinated terrorist attack hits London

(July 7) British Prime Minister Tony Blair speaks about London bombings

(July 7) Bomb scare closes main Edinburgh thoroughfare

(July 7) France raises security level to red in response to London bombings

(July 6) Tanzania accused of supporting terrorism to destabilise Burundi

5.4 Failure analysis Evaluation of the methods described here is non-trivial. We found that our peak detection method is effective despite its simplicity. Anecdotal evidence suggests that our approach to finding explanations underlying unusual spikes and drops in mood levels is effective. We expect that it will break down, however, in case the underlying cause is not news related but, for instance, related to celebrations or public holidays; news sources are unlikely to cover these.

6 Conclusions

We described a method for discovering irregularities in temporal mood patterns appearing in a large corpus of blog posts, and labeling them with a natural language explanation. Our method shows that simple techniques based on comparing corpus frequencies, coupled with large quantities of data, are effective for identifying the events underlying changes in global moods.

References

T. Dunning. 1993. Accurate methods for the statistics of surprise and coincidence. *Comput. Ling.*, 19(1):61–74.

J. Kleinberg. 2002. Bursty and hierarchical structure in streams. In *Proc. 8th ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining*, pages 1–25.

R. Kumar, J. Novak, P. Raghavan, and A. Tomkins. 2003. On the bursty evolution of blogspace. In *Proc. 12th Intern. World Wide Web Conf.*, pages 568–576.

G. Mishne and M. de Rijke. 2006a. Capturing global mood levels using blog posts. In *AAAI 2006 Spring Symp. on Computational Approaches to Analysing Weblogs (AAAI-CAAW 2006)*. To appear.

G. Mishne and M. de Rijke. 2006b. MoodViews: Tools for blog mood analysis. In *AAAI 2006 Spring Symp. on Computational Approaches to Analysing Weblogs (AAAI-CAAW 2006)*.

G. Mishne. 2005. Experiments with mood classification in blog posts. In *Style2005 – 1st Workshop on Stylistic Analysis of Text for Information Access, at SIGIR 2005*.

T. Nanno, T. Fujiki, Y. Suzuki, and M. Okumura. 2004. Automatically collecting, monitoring, and mining Japanese weblogs. In *Proc. 13th International World Wide Web Conf.*, pages 320–321.

J.G. Shanahan, Y. Qu, and J. Wiebe, editors. 2005. *Computing Attitude and Affect in Text: Theory and Applications*. Springer.