MoodViews: Tracking and Searching Mood-Annotated Blog Posts

Gilad Mishne   Krisztian Balog   Maarten de Rijke   Breyten Ernsting

ISLA, University of Amsterdam
Kruislaan 403, 1098 SJ Amsterdam
The Netherlands

gilad,kbalog,mdr,bernstin@science.uva.nl

Abstract

We demonstrate the next release of MoodViews, a set online tools mood analysis in blogs. Since its initial launch in mid-2005, MoodViews has provided a window into aggregate states of mind of masses of people. In addition to the tracking functionalities that MoodViews has offered so far, we demonstrate several types of mood-related search tools. These include searching for moods most closely associated with a given topic, and ranking blog posts not just by publication date or relevancy for a topic (as most blog search engines do), but also by mood (e.g., the "happiest" post on a given topic).

1. Introduction

Some blogging platforms (e.g., LiveJournal and MySpace) allow bloggers to assign a mood descriptor to their blog posts. The post is then published with an indicator of the "current mood" of the blogger, at the time of posting the blog. MoodViews [6] is a platform for collecting, analyzing, and displaying aggregate moods in the blogspace. Launched in mid-2005, MoodViews continuously collects these mood indications, as well as the blog posts themselves, and provides a number of services based on these. More specifically, the data collected is all LiveJournal public posts; at the time MoodViews was created, this amounted to about 100,000 posts every day, but the amount has decreased since to around 50,000 daily posts. In total, in December 2006, MoodViews provides access to more than 35 million blog posts.

MoodViews has been used by several hundreds of thousands of unique visitors and referenced in many thousands of blog posts, as well as in mainstream media sources.

We describe the tools that originally made up MoodViews, as well as the tools that will be part of the next release of MoodViews, scheduled for early 2007.

2. Tracking moods

The tools that originally made up MoodViews—Moodgrapher, Moodteller, and Moodsignals—were aimed at allowing users to track and compare mood levels, i.e., fractions of blog posts tagged with a certain mood.

Moodgrapher. This tool displays the aggregate counts of the different moods in the LiveJournal data, plotting them over
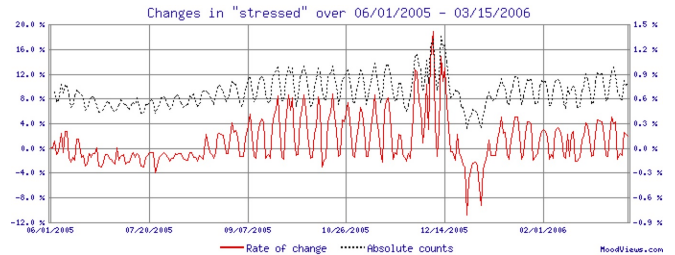
**Fig. 1:** *Levels of* stressed*, June 1, 2005–March 15, 2006*

time. This gives viewers an indication of current trends of moods among bloggers, and allows for various kinds of observation. E.g., some moods display a cyclical behavior, with daily, weekly, or yearly cycles [1], while large scale events (terrorism, natural disasters, global cultural events) leave a clear mark on the aggregate moods of bloggers worldwide. Figure 1 shows the levels for the mood *stressed* (dashed black curve) together with the rate of change for *stressed* over a 9.5 month period; weekly swings are clearly visible, as are seasonal trends (drops during the summer, a peak towards the end of year, followed by a steep drop). See [5] for further examples.

Moodteller. This is an online implementation of a regression-based mood level prediction method described in [4]; it continuously estimates mood levels in the blogspace, shows plots of these estimates, and compares them to the actual, reported mood levels.

Moodsignals. Finally, Moodsignals detects irregular patterns of mood levels, and attempts to provide a natural-language explanation for them; using frequency comparisons and simple burstiness models, it detects words and phrases that overly used during the time of an irregular mood swing, and then uses these to consult contemporaneous news messages; the best matching news messages are then returned as an explanation for the irregular mood pattern [2].

3. Search by mood

In the next release of MoodViews we will add three types of search and browse functionalities (on top of the mood tracking-oriented facilities), allowing users to move from overview/tracking plots for a specific mood to the blog posts on which the plots are ultimately based.
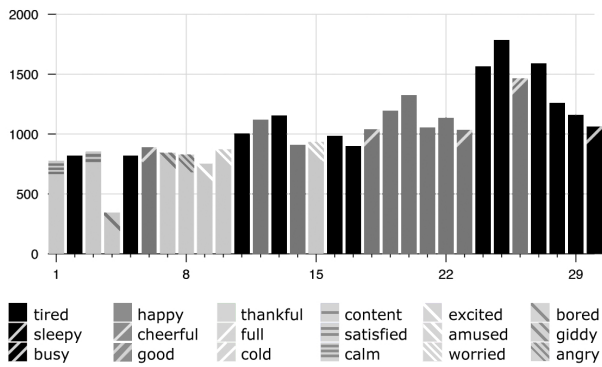
Fig. 2: *(Top): Moodspotter output for "shopping" during November 2006. (Bottom): Color coding for the colors used in the Moodspotter plot*

Moodrecipes. As part of Moodteller, we identify the most distinguishing words and phrases used in posts tagged with a given mood; the frequencies of these most distinguishing words and phrases are used as features in Moodteller's prediction process. Both the words and phrases in the list, and their ranking, are adjusted several several times a day to reflect the language usage in recent posts.

The most distinguishing words and phrases for a given mood are identified as follows. Specifically, for each mood m we define two probability distributions, Θ_m and $\Theta_{\overline{m}}$, to be the distribution of all words in the combined text of blog posts tagged with mood m, and the distribution of all words other posts, respectively. We then rank all the words in m according to their log likelihood measure [7], as compared with m, yielding a ranked list of "characteristic terms" for m.

For every mood, the *mood recipe* for that mood is simply the list of top 10 most characteristic words and phrases used in blog posts tagged with that mood. As an new feature in Moodgrapher and Moodteller, we show the mood recipe for the mood being displaying. The words and phrases are ordered by the degree to which they are distinguishing, and they are clickable: a click takes the user to blog posts (tagged with the mood being displayed) that contain the word or phrase. The idea behind Moodrecipes is that they capture the typical topics associated with a given mood.[1]

Moodspotter. The functionality implemented in Moodspotter is this: given a topic of interest, Moodspotter identifies the moods that "typically" or "most dominantly" associated with the topic, over some period of time. For efficiency reasons, we limit the period being displayed by Moodspotter to one month; Figure 2 shows example output of Moodspotter: the typical moods for the topic "shopping" during November 2006 (a column corresponds to a day).

Our baseline implementation of Moodspotter simply takes the "typical" or "most dominant" mood associated with a topic to be the mood that is the most frequent one in the posts relevant to the topic (on a given day). This may lead to suboptimal results, as this sometimes turns out to be the most frequent mood irrespective of the topic. As an alternative, we phrase the search for the typical mood associated

---

[1] For a detailed case study into the "reasons for happiness" in the blogspace, see [3].

with a topic in probabilistic terms: what is the probability of a mood m being associated with the query topic q? Using simple language modeling techniques, we rank the collection of blog posts according to their relevance for the topic; we then examine each post and if relevant, we examine the mood label of this post.

Rank by moodiness. Using technologies developed for Moodteller and Moodrecipes, the new version of MoodViews will also provide a "rank by mood" facility for restricting search in our indexed blog posts. Users can restrict their search to a specific mood m, and then decide to rank the results either by recency or relevance (like most blog search engines) or by "moodiness," that is, by the degree to which the (language used in the) relevant post is typical for the mood m.

4. Conclusions

In this paper, we described the next release of MoodViews. In addition to the mood tracking tools available in the first release, we are expanding MoodViews with search and browsing facilities.

Acknowledgments

References

[1] K. Balog and M. de Rijke. Decomposing bloggers' moods. In *WWW-2006 Workshop on the Weblogging Ecosystem*, 2006.

[2] K. Balog, G. Mishne, and M. de Rijke. Why are they excited? identifying and explaining spikes in blog mood levels. In *Proceedings 11th Meeting of the European Chapter of the Association for Computational Linguistics (EACL 2006)*, April 2006.

[3] R. Mihalcea and H. Liu. A corpus-based approach to finding happiness. In *AAAI Spring Symposium on Computational Approaches to Analysing Weblogs*, 2006.

[4] G. Mishne and M. de Rijke. Capturing global mood levels using blog posts. In N. Nicolov, F. Salvetti, M. Liberman, and J. Martin, editors, *AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW 2006)*, pages 145–152. AAAI Press, March 2006.

[5] G. Mishne and M. de Rijke. MoodViews: Tools for blog mood analysis. In N. Nicolov, F. Salvetti, M. Liberman, and J. Martin, editors, *AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW 2006)*, pages 153–154. AAAI Press, March 2006.

[6] MoodViews. Tools for blog mood analysis, 2006. URL: http://moodviews.com.

[7] P. Rayson and R. Garside. Comparing corpora using frequency profiling. In *Workshop on Comparing Corpora, held in conjunction with the 38th annual meeting of the Association for Computational Linguistics (ACL 2000)*, 2000.