Understanding How Bloggers Feel: Recognizing Affect in Blog Posts

Gilly Leshed, Joseph ‘Jofish’ Kaye
Cornell University, Information Science Program
301 College Ave., Ithaca NY 14850, USA
gl87 | jofish at cornell.edu

ABSTRACT
One of the goals of affective computing is to recognize human emotions. In this paper, we present a system that learns to recognize emotions based on textual resources. We use as input a large number of blog entries tagged by their authors with moods during the course of writing, rather than post-hoc or third-party labeling. We show how words in the texts exemplify the moods, present satisfactory results for the mood recognition task, and illustrate a method for analyzing mood synonymy within blog entries. Our findings suggest that an automated machine-learning approach can be used to gain insight into the way writers convey and interpret their own emotions.

Author Keywords
Affective computing, blogs, emotions, natural language processing, support vector machines, text categorization.

ACM Classification Keywords
H.3.1 [Information Storage And Retrieval]: Content Analysis and Indexing

INTRODUCTION
Automated recognition of human emotion has long been a goal of affective computing. Affect recognition can not only give us a richer understanding of cognitive processes within the human mind, but assist us in creating computers that better react to human input. Existing research on affect recognition has mainly looked at sources of facial expressions [3], vocal intonation and physiological signals that vary with affective states [10]. However, although language is an important way to convey emotions, either in oral forms such as public speeches or in written forms such as prose or poetry [9], little work has been conducted on detecting emotions based on verbal expressions. In this paper, we demonstrate an automated process which can help better understand how writers feel.

Blogs as a Resource for Affective Texts
A blog is a web application that contains periodic posts, also called blog entries, on a common webpage. The vast majority of blogs is written for personal communication and expression and can be viewed as personal diaries, where bloggers write about their experiences, opinions, and emotions. As a result, blogs provide naturally occurring windows into people’s thoughts and feelings [1].

LiveJournal has over eight million users, of whom 1.5 million have updated their blog in the last month (www.livejournal.com/stats/). One of the features of LiveJournal is that it allows users to tag their posts with two particular types of tags: a mood tag and a music tag as shown in Figure 1. The large number of mood-tagged entries presented us with a particular opportunity: to attempt to build a system that can recognize human moods.

LiveJournal bloggers can choose from 132 predefined moods, including happy, sad, and confused, but also bitchy, blah and indescribable. They can also type in any other mood, or leave this field blank. These moods are clearly not, for instance, the six ‘basic’ categories of emotion enumerated by Ekman [3]. In fact, the wide variety of moods used by bloggers may support research questioning the notion of a ‘basic emotion’ [7]. We deliberately decided to base our choices of mood categories only on their actual use by bloggers. In this way, we aspired to understand how bloggers interpret their moods, rather than presuppose any notion of whether a given term is or is not an emotion.

Figure 1. A LiveJournal blog entry with attached mood.
Affective Text Analysis Approaches
Given a large quantity of tagged entries, how can an effective system automatically recognize emotion within them? Existing techniques fall into two primary approaches: linguistic analysis and machine learning text categorization.

Linguistic Analysis
Linguistic analysis approach for emotions attempts to understand peoples’ psychological and social states according to linguistic characteristics of their verbal expressions, either spoken or written. Research shows it is possible to identify linguistic indicators for emotional states of writers. For example, Cohn et al. measured the emotional sense of blogs before and after the September 11th attacks, using a database of pre-determined word senses. They found that post-9/11 blog entries showed more negative emotions and were more cognitively and socially engaged, even among bloggers who hardly wrote about the events themselves [1]. Linguistic analysis approaches rely on the researcher’s beliefs and interpretations regarding linguistic representation. These may not correspond to the authors’ conscious or unconscious intentions. Additionally, constructing the feature set can be costly in terms of human labor [12].

Automated Text Categorization
A second approach that can be applied for predicting affect verbal expressions is automated text categorization. In this approach texts are classified into a predefined set of categories, such as emotions. In other words, for each document, we attempt to identify, out of a set of possible emotions, the one that most closely characterizes the document text.

The leading approach in automated text categorization at present involves various machine learning techniques [12]. This approach involves an inductive process, also called the learner, which observes a formerly classified set of documents, the training set, and assembles the typical characteristics of classified documents. This assembly is the basis for the classifier, which in turn is fed by new documents, the test set, and does the actual text categorization. Among a large variety of machine learning techniques for text categorization, we chose to use Support Vector Machines [4] for our study. Studies have shown that SVMs perform best among several machine learning methodologies in experimental settings using large training sets [4,8].

An alternative approach for machine learning involves knowledge engineering strategies. Liu et al, for instance, used a commonsense knowledge base, to tag sentences with basic emotions [5]. While they generated an application felt to be intelligent and interactive by users, their method, much like the linguistic analysis approach, relies upon a labor-intensive hand-crafted repository.

METHOD
We randomly selected 100,000 LiveJournal blogs and downloaded their RSS feeds. These blogs contained 1.4 million blog entries, with about 10% containing no posts at all, about 10% containing only one post, and about 42% containing 25 posts, the maximum number of posts captured by the RSS-feeds. In about 58% of the 1.4M blog entries bloggers chose to attach moods to the entries, giving us a total dataset of approximately 812,000 mood-tagged blog entries of average length 168 words. (This dataset is available on request from the authors.)

In order to determine how many moods need to be categorized, we counted the number of unique moods. The ten most frequently used moods are, in descending order of frequency: tired, amused, happy, bored, blah, cheerful, content, sleepy, excited, and calm. We collected examples of approximately 100,000 unique moods from our sample, although only the top 140 moods were used more than a hundred times. To ensure we had sufficient data to apply the SVM, we only analyzed blog entries tagged with one of the top fifty moods.

To create input for the algorithm, each blog entry was indexed into a feature vector using a standard tf/idf scheme [11]. This information retrieval technique treats a document as a “bag of words”, assuming that the presence of words is significant, but their order is not, in contrast to, for example, [5,6]. As is standard practice in information retrieval tasks, not all words in the blog entries were indexed, but only the 5,000 most frequent ones, excluding a standard list of stopwords, like ‘and’ or ‘is’. Too rare or frequent words do not assist in discriminating between categories, and thus ignoring them reduces computational needs.

We applied an SVM learner, SVMlight (svmlight.joachims.org) on the blog entry feature vectors of the training set. The results were used by the SVMlight classifier to classify the feature vectors of the test set. This binary classifier looks at each mood category separately, and decides if each blog entry belongs to that mood or not. We then compared SVM’s decision with the actual mood of the entry to evaluate the performance of the algorithm.

RESULTS

Word and Mood Associations

Words Most Relevant for Moods
Our first result is a list of words that characterizes each mood. In Table 1, we show the ten words rated by the SVM learner as most characteristic for a sample of moods. For instance, a blog entry tagged with the mood loved typically included the words valentine, sweetest, and roses.

<table>
<thead>
<tr>
<th>mood</th>
<th>associated word</th>
</tr>
</thead>
<tbody>
<tr>
<td>sad</td>
<td>valentine</td>
</tr>
<tr>
<td>died</td>
<td>sweetest</td>
</tr>
<tr>
<td>crying</td>
<td>loved</td>
</tr>
<tr>
<td>sadness</td>
<td>roses</td>
</tr>
<tr>
<td>cry</td>
<td>kissed</td>
</tr>
<tr>
<td>upset</td>
<td>arms</td>
</tr>
<tr>
<td>funeral</td>
<td>necklace</td>
</tr>
<tr>
<td>passed</td>
<td>eachother</td>
</tr>
<tr>
<td>memories</td>
<td>layed</td>
</tr>
<tr>
<td>cried</td>
<td>hugged</td>
</tr>
</tbody>
</table>

Table 1. Five examples of words most associated with moods.
Topic-based text categorization systems often yield performance measures of around 80% [12]. Yet, the nature of blogs and the emotions we are trying to predict introduces bias into the task. For instance, having the bloggers themselves select the mood for their blog entry makes the selection subject to their understanding of the meaning of a mood. We see this as a trade off: the advantage of ecological validity of self-reported mood in the course of writing over the consistency of predetermined emotional categories.

By contrast, a random guess, given the distribution of positive and negative examples we chose, would have yielded an accuracy level of 25%. Our results suggest that using a large corpus is indeed powerful in achieving superior results, even with the simple bag-of-words approach: [6] applied a complex set of mood recognition cues on a dataset from the same corpus and produced lower accuracy rates, even on datasets of comparable size. This is a difficult task: even humans would not be accurate in consistently distinguishing between, say, annoyed and pissed off.

Sentiment Classification: Positive vs. Negative Affect
To verify whether the large amount of moods accounted for the performance values, we manually sorted the moods into positive and negative classes, excluding moods that could not be straightforwardly classified into either class. As expected, the system was better at this task than the multiple mood classification, and had a 74% accuracy rate, a 72% precision rate, and an 80% recall rate, comparable to other sentiment classification tasks [8]. The findings suggest that a simple method is promising for applications in affective computing in which it is sufficient to know whether the user is in a good or bad mood.

Mood Synonymy
Over the years, researchers have proposed that there exist between two and twenty basic emotional categories [10], although others disagree [7]. The relevant words found by the SVM, discussed earlier, suggest that some moods may be synonymous, and fall within the same mood category. For example, it seems that the moods angry and aggrivated are synonymous, both highly weighing the words anger, bastard, and assholes in their weighted word lists.

To estimate possible similarity between moods, we computed the cosine similarity between the weight vectors created by the SVM learner, according to the formula:

$$\cos(sim(u, v)) = \frac{\sum_{i=1}^{n} (w_{ui} \times w_{vi})}{\sqrt{\sum_{i=1}^{n} w_{ui}^2} \times \sqrt{\sum_{i=1}^{n} w_{vi}^2}}$$

where $u$ and $v$ represent two moods, and $w_{ij}$ represents the weight of word $i$ in the weight vector of mood $j$. Positive values of this measure indicate similarity between the two moods, and respectively, negative values indicate dissimilarity between the moods. The resulting similarity measures between all dyad moods were then clustered into ten groups using hierarchical clustering analysis, shown in Figure 3.
Each cluster represents a group of moods that are synonymous on the level of the hierarchy cut. Some clusters are composed of a single mood, implying that these moods were distinctly different from any other mood.

Some of the clusters are clearly clusters of positive or negative moods. However, while it is easy to understand how sleepy and tired are in the same cluster, it is surprising to find awake, creative, and amused together with them. The division into clusters does not necessarily match our intuitions about basic emotion categories. However, the mood clusters are rooted in bloggers’ expression of themselves. The clustering technique is not subject to anyone’s preconceived judgment: it simply groups moods that share similar patterns of word use by bloggers.

CONCLUSIONS
The results show that, to some degree, we can predict emotional states of bloggers from their writings. Affective computing to date has often relied on classifying objective measures, such as galvanic skin response or respiration rate, into a set of emotions predetermined by the researcher. In contrast, our use of self-reported moods generated in the course of everyday life emphasizes the lived experience of mood by the writer. This can be a limitation: when bloggers select a mood, their purpose is not to correctly identify their emotional state and correlate it with the entry they have written. However, our testing strategy compares the automated prediction to the users’ perceptions of their own mood.

A key element is the harvesting and use of a very large pre-existing corpus generated by many users. Even recognizing the existence of group blogs and multiple blogs written by single individuals, our N is still on the order of tens of thousands of users. By gathering examples from natural settings, we have an claim for ecological validity absent from many laboratory studies of emotion.

Our method does have limitations: the bag-of-words approach ignores negation constructions (“not sad”), and we conflate all bloggers into a single dataset. Our results may not apply outside of the limited domain of blogging. We include moods that do not represent traditional emotions (hungry, sleepy). Our accuracy levels on the 50-mood set are low compared to other text classification tasks, probably because the mood set is very large relative to other studies, i.e. [3,8]. The results obtained for the positive versus negative mood classification emphasize that smaller class sets may lead better performance. Our work on mood synonymy is a first step in applying this conclusion, and our current work involves identifying entries as one of these meta-categories.

ACKNOWLEDGMENTS
We thank Claire Cardie, Paul Houle, Thorsten Joachims, Moshe Koppel, Eugene Medynskiy, Roz Picard, Carson Reynolds, and Phoebe Sengers for their advice.

REFERENCES