Finding Emotion in Image Descriptions

Morgan Ulinski  
Columbia University  
New York, NY, USA  
mulinski@cs.columbia.edu

Victor Soto  
Columbia University  
New York, NY, USA  
vs2411@columbia.edu

Julia Hirschberg  
Columbia University  
New York, NY, USA  
julia@cs.columbia.edu

ABSTRACT

In this paper, we approach the problem of classifying emotion in image descriptions. A method is proposed to perform 6-way emotion classification and is tested against two labeled datasets: a corpus of blog posts mined from LiveJournal and a corpus of descriptive texts of computer generated scenes. We perform feature selection using the mRMR technique and then use a multi-class linear predictor to classify posts among the Ekman Big Six emotions (happiness, sadness, anger, surprise, fear, and disgust) [9]. We find that TFIDF scores on lexical features and LIWC scores are much more helpful in emotion classification than using scores calculated from existing sentiment dictionaries, and that our proposed method performs significantly better than a baseline classifier that chooses the majority class. On the blog posts, we achieve 40% accuracy, and on the corpus of image descriptions, we achieve up to 63% accuracy.

Categories and Subject Descriptors

I.2.7 [Artificial Intelligence]: Natural Language Processing—Text Analysis; I.5.4 [Pattern Recognition]: Applications—Text Processing

General Terms

Algorithms, Design, Experimentation

Keywords

image descriptions, mood classification, text tagging

1. INTRODUCTION

Detecting emotion from text has been an increasingly popular research topic in recent years. Applications for automatic emotion detection range from advertisement and commercial purposes to medical patient behavior analysis. Applied in a social network setting, emotion detection can be a powerful tool to gain knowledge about how individuals, social circles, communities, or cities feel about current events or other topics.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. WSDM’12, August 12 2012, Beijing, China.  
Copyright 2012 ACM 978-1-4503-1543-2/12/08 ...$15.00.

The problem of emotion detection poses interesting questions from a research point of view; for instance: how to model the text for the detection task, what features offer the best prediction/detection power, and to what extent it is even possible to accurately distinguish subjective labels such as emotions from a given source text.

Emotion detection has been studied on a variety of types of corpora. In our work, we focus on a genre of text that has not been considered as much as some of the others with respect to emotion detection; namely, captions and comments on images. Descriptions and captions on images appear in many contexts, especially in social media. Facebook users post pictures on their walls, adding a caption or description to the image. Twitter posts can include a similar caption with a link to an image. Blog posts will often have an image as the main content for the post, with a small caption or description attached. In our work, we will focus on a different context for image captions: the image descriptions used as input to a text-to-scene program. This has practical value in that we will be able to extend a text-to-scene program to include not just literal translations of a description, but also emotional elements that can show up in the resulting image’s lighting, color scheme, etc.

To predict emotion, we carry out a fairly traditional machine learning method with the addition of feature selection techniques. Specifically, the experiments here use a set of six basic emotions: happiness, sadness, anger, surprise, fear, and disgust. These are the Ekman Big Six emotions. Two datasets are used in the experiments. The first dataset consists of a collection of 300,000 user-annotated blog posts, extracted from LiveJournal (http://www.livejournal.com). The second one consists of a collection of captions and comments of computer-generated images from WordsEye (http://www.wordseye.com). A model is built for each post with more than 30,000 features extracted from six different sources, including frequency features, syntactic features, lexical features, and sentiment scores. Afterwards, the instance space is reduced by applying feature selection techniques. Finally the classification is performed using LIBLINEAR [11], a library of linear support vector machines for large and sparse datasets.

2. STATE OF THE ART

Recent research on NLP has started to utilize user-generated content like blog posts in areas like Opinion Mining and Sentiment Analysis. For instance, the authors of [1] and [4] use political blogs. They measure disagreement among users and polarity of comments, respectively. In [2], a sentiment analysis study is performed over a Twitter dataset. However, research on emotion classification is still relatively scarce compared to sentiment tasks. In [14], Mishne used LiveJournal and SVMs to perform binary classification over a large set of emotions, with results slightly better than the 50% baseline. In [16], the authors perform binary classifi-
cation on the Big Six emotions using a variety of classifiers over a
LiveJournal dataset. The work in [8] proposes a word-to-sentence
emotion classification scheme. Using SentiWordNet features and
emotion indicators at a word level, the experiments over a Bengali
blog corpus yield mood accuracies between 60 – 72%. Finally, the
authors in [5] perform multiclass classification using SVM over the
UIUC children’s stories corpus. Their best accuracy on an emotion
different than Neutral is 23%. Our work differs from these pre-
vious efforts in that we perform multi-class classification over our
emotion set on the texts, and in the feature selection techniques we
perform prior to any classification. Comparing the results among
these studies and establishing a baseline is hard because often ei-
ther the data source, the set of emotions, or the computed features
differ. Even when the data source coincides, the examples in the
dataset are not the same. This fact highlights the immature state of
research area and the need for a standardized dataset for emo-
tion classification.

3. THE DATASETS

For our purposes we will test the proposed methodology in two
different datasets. The first dataset is large, sparse and very noisy.
The second dataset is smaller, its dictionary more restricted, and
the set of moods better defined.

3.1 The LiveJournal Dataset

The initial dataset consisted of about 300,000 labeled blog en-
tries downloaded from LiveJournal. The dataset was collected by
scrapping Livejournal for entries that had been tagged with a mood.
Ultimately, we ended up downloading blogs from about 30,000
users. Unfortunately, the distribution of the six basic emotions we
are interested in across the entries was quite uneven. The primary
problem is that when users select a mood with which to label a blog
post, they are provided with a drop-down list of about 132 moods.
Five of the six basic emotions: happy, sad, angry, surprised and
afraid, appear explicitly in the list in adjective form. The sixth,
disgusted, does not; thus, in order to use disgusted as a mood tag
for their posts, users must manually type in the word. This results
in a dearth of posts labeled as disgusted as compared to the other
more common moods. To counter this problem, we also included
posts tagged with reasonable synonyms of the six basic emotions
in our dataset for those moods with fewer posts available. These
synonyms are listed in Table 1. Any entry tagged with a mood
not contained in this table was discarded. Three consecutive filters
were applied to the dataset. First, with the goal of reducing the size
of the dataset and maintaining a consistent vocabulary among the
entries, we decided to use only those entries written in November
and December. By limiting to some extent the possible contexts
of the blog posts (that is, we expected many entries written during
these months to be related to the holidays, or to winter, for exam-
ple), we hoped to minimize some of the variation in our dataset.
Second, any entry not written in English was discarded and, third, 
any entry with an empty body of text, even if it had a title, was
discarded as well.

3.2 The WordsEye Dataset

WordsEye [7] is an automatic text-to-scene system that allows
one to create 3D scenes using natural language. The system has a
large database of 3D models whose features can be edited, along
with their relative position to other objects in the scene, textures,
lighting, etc. One desired future expansion of WordsEye is the
ability to set the mood of the scene automatically, for example, by
manipulating its lighting or the predominant colors. The WordsEye
website allows the author of a picture to write a title and a caption.

<table>
<thead>
<tr>
<th>Class</th>
<th>Labels</th>
<th># blog entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>happy</td>
<td>1257</td>
</tr>
<tr>
<td></td>
<td>sad</td>
<td>639</td>
</tr>
<tr>
<td>Anger</td>
<td>infuriated, pissed off, enraged</td>
<td>681</td>
</tr>
<tr>
<td></td>
<td>imitated</td>
<td></td>
</tr>
<tr>
<td>Surprise</td>
<td>surprised, amazed, shocked</td>
<td>432</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fear</td>
<td>scared, afraid, worried, anxious, nervous</td>
<td>1002</td>
</tr>
<tr>
<td>Disgust</td>
<td>disgusted, displeased, fed up, repulsed</td>
<td>432</td>
</tr>
<tr>
<td></td>
<td>revolted, scandalized, sick and tired, turned off, ew, yuck</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: LiveJournal labels and examples per class

Please write a literal description (3-5 sentences) that could
work as a caption for the previous image. Avoid the use of emo-
tional vocabulary (min 200 characters)

This is a picture of a man in a purple suit. He is wearing blue sun-
glasses and is sitting on an orange couch. There is a lamp and a
table in the room and the wall is made of brick. There is also a grey
carpet in the room.

Please choose the mood that relates the most to the displayed
picture:

Anger

Why did you choose that mood? Write a short answer (min 60
characters). This time emotional vocabulary is allowed. Make
sure the answer is significantly different from the description,
or the HIT will not be approved.

The man in the picture looks angry. His eyebrows make him look
that way, due to the way they are slanted.

Table 2: Real completed task
It also enables the rest of the users to comment on the picture and their impressions about it. Our goal is to be able to accurately set a mood for the scene based on this kind of data.

In order to create this kind of dataset, we collected a total of 660 pictures generated by previous users of the WordsEye system. Each picture was labeled using the Amazon Mechanical Turk, a crowdsourcing marketplace for human intelligence tasks (https://www.mturk.com/). Specifically, the designated task consisted of the following parts: (a) a WordsEye picture was presented to the worker, (b) the worker was asked to provide a literal description that could work as a caption for the image, explicitly avoiding the use of emotional words (the minimum length of each caption was set to 200 characters), (c) the worker tagged the picture with the mood that relates the most to the scene (out of the six big moods) and finally (d) the users were asked to comment on the reasons for which they selected that mood, this time allowing the use of emotional words (the minimum length for this question was set to 60 characters). Tables 2, 3, and 4 show some completed tasks.

For each completed task, the worker was rewarded $0.30. In total, each picture underwent this procedure three times and therefore was assigned three different captions, comments and mood tags. It would be possible to hone down the dataset by keeping only those images that were associated with an unambiguous mood; that is, images on which the three turers chose identical moods. However, we are interested in subjective and personal opinions, and thus preferred to keep all individual judgements. Since our classification is based only on the descriptive text and not on the source image, it is acceptable to consider each description as a separate datapoint. Table 5 shows the distribution of mood labels in the dataset, which is again highly unbalanced. For each image, a document was created that contained both the caption and the comments. Every reference to the mood in the comments was removed and substituted by the tag <mood> to avoid introducing a bias with the class labels.

4. FEATURES

For each example in a dataset a feature vector is built containing the following information:

(a) The class label.
(b) SentiWordNet scores [10, 3]. SentiWordNet is a lexical tool for opinion mining. It assigns three scores to every WordNet [13, 12] synset: a positivity score, a negativity score and an objectivity score. The SentiWordNet scores of a document are computed via the mean scores of every word in the document.
(c) LIWC scores [17]. LIWC is a text analysis tool that calculates 81 language features in several categories including general descriptors, linguistic dimensions, psychological constructs, personal concerns, paralinguistic dimensions and punctuation. For instance, to compute the posemo feature, LIWC uses a list of 406 terms that includes words like love, nice, and sweet, and to compute sad a list of 101 words is used which comprises terms like crying and grief.
(d) Dictionary of Affect scores [18]. The Dictionary of Affect is a lexical tool to measure the emotional meaning of texts. The DAL assigns activation, evaluation and imagery scores to every word. It does so by comparing each word to a list of 8700+ words rated by their activation, evaluation and imagery. The DAL score of document is given by the average values of its word scores.
(e) TF-IDF of word-POS tag pair. For each document in the corpus, the TF-IDF score of each possible pair formed by a n-gram word stem and its n-gram part-of-speech tag is computed, with \( n = \{1, 2, 3\} \). Approximately, the LiveJournal dataset dictionary contains 27500, 210000 and 14000 1-grams, 2-grams and 3-grams respectively, whereas the WordsEye dataset contains 3500, 13000 and 20000 of each.
In this section the methodology followed in our experiments is explained. After building the model of each blog entry, the full dataset contained 4505 vectors comprised of a class label and 34,312 features. Besides the obvious high dimensionality of the problem, the set is still unbalanced, as seen in Table 1. Starting from the original dataset, 10 balanced datasets are created by sampling the original set without replacement. The balanced datasets have 2520 examples each and 420 instances per label. The prediction estimates over each of these balanced datasets are averaged to eliminate possible bias derived from the random selection of instances. Over each balanced dataset, 10-fold cross validation is performed. In total, for each balanced dataset 10 new datasets are created where the training set contains 90% of the examples and the test set contains 10%. Feature selection techniques will be applied to the training set of each balanced dataset and returns the best 500 features for the prediction problem. The top twenty features are shown in Table 6. Features in bold font are LIWC features; the rest are TF-IDF scores. The ranking is computed by the average min-redundancy max-relevance score of each feature over each partition and fold.

For the LiveJournal and full WordsEye dataset, most of the best selected features are either TF-IDF scores (represented by lemma1 . . . lemmaN, posTag1 . . . posTagN) or LIWC features (in bold fonts), which suggests that the sentiment scores computed with SentiWordNet (whose features, except positive score, appear past the nineteenth position) and DAL (whose features appear after the four-hundredth position) are not very useful for our classification problem. It would be interesting to explore the use of more detailed emotion dictionaries that provide scores for individual emotions rather than just polarity, as these would likely be more useful for our task. Most of the features selected by the mRMR technique seem to be consistent with the task at hand, referring to words and concepts intuitively associated with emotions.

### 7. Experiments and Results

The library LIBLINEAR [11] was used to perform all the classification experiments presented in this work (experiments were also carried out using multiclass SVMs from LibSVM with gaussian kernels but the results were indistinguishable and the runtime at least 10 times longer). We tried binary classification of each emotion, but the results were not very promising. Instead, we focused on direct multiway classification. The precision, recall and F-score results of the 6-way SVM classification using the features returned by the mRMR method are shown in figures 2, 3 and 4. The accuracy plots are shown in figure 1. Finally, table 7 shows the average statistics for the number of features that maximize the average accuracy. Because the datasets are balanced, the baseline accuracy is 1/6 = 16.7%. Our method achieves vastly different results for each dataset. An average maximum accuracy of 40% at 128 features is reached over the LiveJournal dataset, which is significantly higher than the baseline. The WordsEye dataset of captions only yields the worst results, which is to be expected given the short length of the texts and the lack of emotion-related words in the dictionary. Even so, it slightly improved over the baseline at 26% accuracy using 96 features. The best results were achieved with the full WordsEye dataset, peaking at 62.5% accuracy using 16 features.

Our method obtains good precision results over the second WordsEye dataset, but fails at achieving good recall values with the Disgusted and Surprise classes, which ultimately causes a drop in the F-Score and accuracy values. This is most apparent in table 8, which shows an average confusion matrix from the sub-experiments that achieved 66% accuracy. Each row (from top to bottom) corresponds to the instances labeled as happy, sad, angry, surprised, scared and disgusted and each column from (left to right) corresponds to the instances predicted as the same classes in the same order. The confusion matrix highlights how Disgusted and Surprise examples are often misclassified. It also shows a certain bias to assign Happiness as the mood when making a mistake. The use of several mood labels for each class on the LiveJournal dataset has direct consequences; both Happiness and Sadness, the only classes
### Table 6: Selected features by mRMR. Bold fonts indicate LIWC feature, (SWN) indicate SentiWordNet feature, and the rest of the features are tfidf features of n-grams

<table>
<thead>
<tr>
<th>Pos.</th>
<th>LiveJournal</th>
<th>Captions</th>
<th>Captions+Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>sad</td>
<td>red,NN</td>
<td>negemo</td>
</tr>
<tr>
<td>2</td>
<td>anger</td>
<td>of the, IN DT</td>
<td>&lt;mood&gt;, JJ</td>
</tr>
<tr>
<td>3</td>
<td>posemo</td>
<td>human, JJ</td>
<td>anx</td>
</tr>
<tr>
<td>4</td>
<td>disgusting,JJ</td>
<td>brain, NN</td>
<td>posemo</td>
</tr>
<tr>
<td>5</td>
<td>sad,JJ</td>
<td>dinosaur,NNs</td>
<td>sad</td>
</tr>
<tr>
<td>6</td>
<td>pppron</td>
<td>figure, NNS</td>
<td>anger</td>
</tr>
<tr>
<td>7</td>
<td>die, VBD</td>
<td>tank, NN</td>
<td>gross, JJ</td>
</tr>
<tr>
<td>8</td>
<td>wanna, VB</td>
<td>while the, IN DT</td>
<td>add, JJ</td>
</tr>
<tr>
<td>9</td>
<td>anx</td>
<td>there be several, EX VBP JJ</td>
<td>mad, JJ</td>
</tr>
<tr>
<td>10</td>
<td>death</td>
<td>picture there be, NN EX VBZ</td>
<td>scary, JJ</td>
</tr>
<tr>
<td>11</td>
<td>. she be, SENT PP VBD</td>
<td>while, IN</td>
<td>it be very, PP VBZ RB</td>
</tr>
<tr>
<td>12</td>
<td>smile emoticon</td>
<td>hockey mask, NN NN</td>
<td>creep, JJ</td>
</tr>
<tr>
<td>13</td>
<td>shock emoticon</td>
<td>anger</td>
<td>picture because it, NN IN PP</td>
</tr>
<tr>
<td>14</td>
<td>happy, JJ</td>
<td>stand, VBP</td>
<td>posScore (SWN)</td>
</tr>
<tr>
<td>15</td>
<td>incl</td>
<td>dinosaur be, NNS VBP</td>
<td>lonely, JJ</td>
</tr>
<tr>
<td>16</td>
<td>shee</td>
<td>hold, VBG</td>
<td>&lt;mood&gt; because, JJ IN</td>
</tr>
<tr>
<td>17</td>
<td>laugh emoticon</td>
<td>white hockey mask, JJ NN NN</td>
<td>depressing, JJ</td>
</tr>
<tr>
<td>18</td>
<td>western, NP</td>
<td>this, DT</td>
<td>strange, JJ</td>
</tr>
<tr>
<td>19</td>
<td>negemo</td>
<td>skull, NN</td>
<td>disturbing, JJ</td>
</tr>
<tr>
<td>20</td>
<td>posScore (SWN)</td>
<td>percept</td>
<td>figure, NNS</td>
</tr>
</tbody>
</table>

### Table 7: Mean and standard Precision, Recall and F-score values for each dataset

<table>
<thead>
<tr>
<th>Mood</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>0.46±0.02</td>
<td>0.55±0.02</td>
<td>0.50±0.02</td>
<td>0.26±0.08</td>
<td>0.26±0.1</td>
<td>0.25±0.09</td>
<td>0.59±0.04</td>
<td>0.81±0.06</td>
<td>0.67±0.04</td>
</tr>
<tr>
<td>Sad</td>
<td>0.49±0.02</td>
<td>0.49±0.02</td>
<td>0.49±0.02</td>
<td>0.21±0.09</td>
<td>0.22±0.07</td>
<td>0.21±0.07</td>
<td>0.73±0.04</td>
<td>0.77±0.05</td>
<td>0.74±0.04</td>
</tr>
<tr>
<td>Angry</td>
<td>0.38±0.02</td>
<td>0.37±0.04</td>
<td>0.37±0.03</td>
<td>0.33±0.04</td>
<td>0.36±0.05</td>
<td>0.34±0.05</td>
<td>0.62±0.05</td>
<td>0.65±0.05</td>
<td>0.61±0.03</td>
</tr>
<tr>
<td>Surprised</td>
<td>0.32±0.02</td>
<td>0.29±0.02</td>
<td>0.30±0.02</td>
<td>0.20±0.06</td>
<td>0.18±0.08</td>
<td>0.18±0.06</td>
<td>0.64±0.09</td>
<td>0.44±0.09</td>
<td>0.50±0.08</td>
</tr>
<tr>
<td>Scared</td>
<td>0.38±0.01</td>
<td>0.37±0.03</td>
<td>0.37±0.02</td>
<td>0.26±0.08</td>
<td>0.24±0.07</td>
<td>0.24±0.07</td>
<td>0.68±0.05</td>
<td>0.73±0.06</td>
<td>0.69±0.04</td>
</tr>
<tr>
<td>Disgusted</td>
<td>0.38±0.02</td>
<td>0.35±0.03</td>
<td>0.36±0.03</td>
<td>0.27±0.04</td>
<td>0.28±0.05</td>
<td>0.26±0.04</td>
<td>0.63±0.07</td>
<td>0.35±0.09</td>
<td>0.42±0.07</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.40±0.01</td>
<td>0.26±0.02</td>
<td>0.40±0.01</td>
<td>0.26±0.02</td>
<td>0.26±0.02</td>
<td>0.26±0.02</td>
<td>0.63±0.07</td>
<td>0.42±0.07</td>
<td>0.42±0.07</td>
</tr>
</tbody>
</table>

### Figure 2: (Left) Recall, (Center) Precision and (Right) F-score of the 6-class classification problem over the LiveJournal dataset
Figure 3: (Left) Recall, (Center) Precision and (Right) F-score of the 6-class classification problem over the WordsEye caption dataset.

Figure 4: (Left) Recall, (Center) Precision and (Right) F-score of the 6-class classification problem over the WordsEye caption and comment dataset.
defined by a single mood label, yield the best precision/recall results, whereas Surprise and Disgust, whose mood labels contain larger variability (amazed - shocked, disgusted - sickened), have the worst results. Overall the results are superior to the baseline: with only 64 features the precision of the classification for each label ranges from 0.25 to 0.6 and the recall from 0.35 to 0.45.

8. CONCLUSIONS AND FUTURE WORK

We tested a method to perform emotion classification using mRMR feature selection and 6-way SVM classification. Our results indicate that the method is significantly superior to the baseline classifier, using frequency features of pairs of words/POS tags and lexical scores computed with the LIWC software. We also found that the multiway classification was more successful than binary classification of each emotion, although more exploration of that question may be needed. This may be because in the binary classification problem, we have to sacrifice some of our data by downsampling the class of entries not labeled as the mood in question in order to achieve a balanced dataset. In multiway classification, we are able to use all of our data. Upon running the same experiments over two human-annotated datasets, a large corpus of LiveJournal blogs and a smaller corpus of caption and comments attached to computer-generated scenes, we found that our methods yield better results over both the baseline and the latest state-of-the-art.

In future work, we plan to extend the WordsEye dataset with more AMT annotated data. The larger corpus can then be filtered to contain only the collected information of those pictures that received a significantly high number of votes of the same mood. This way, with more consistent labels, we hope to build more accurate and robust classifiers. In addition, we hope to be able to use the large amount of LiveJournal data available to boost our results on the WordsEye data, by incorporating its larger inventory of words. This would make our model more adaptable to new description environments, whereas WordsEye with only 64 features the precision of the classification for each label ranges from 0.25 to 0.6 and the recall from 0.35 to 0.45.

<table>
<thead>
<tr>
<th></th>
<th>H</th>
<th>A</th>
<th>S</th>
<th>E</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>0.83</td>
<td>0.00</td>
<td>0.00</td>
<td>0.17</td>
<td>0.00</td>
</tr>
<tr>
<td>Sadness</td>
<td>0.33</td>
<td>0.67</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Anger</td>
<td>0.17</td>
<td>0.00</td>
<td>0.83</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Surprise</td>
<td>0.33</td>
<td>0.00</td>
<td>0.17</td>
<td>0.50</td>
<td>0.17</td>
</tr>
<tr>
<td>Fear</td>
<td>0.17</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.83</td>
</tr>
<tr>
<td>Disgust</td>
<td>0.00</td>
<td>0.17</td>
<td>0.00</td>
<td>0.33</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Table 8: Average confusion matrix for a WordsEye run (67% accuracy).

9. ACKNOWLEDGMENTS

This research was partially supported by ‘la Caixa’ Fellowship Grant for Post-Graduate Studies, Caixa d’Estalvis i Pensions de Barcelona, and by NSF-IIS-1145505 “EAGER: Using Social Media and Crowdsourcing to Create a New Affect Dictionary”.

10. REFERENCES


