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.

WISDOM'12, August 12 2012, Beijing, China.

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 classification 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 previous 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 either 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 the research area and the need for a standardized dataset for emotion 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 entries downloaded from LiveJournal. The dataset was collected by scraping 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 example), 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.

Class	Labels	# blog entries
Happiness	happy	1257
Sadness	sad	639
Anger	angry infuriated pissed off enraged irate	681
Surprise	surprised amazed impressed shocked	432
Fear	scared afraid worried anxious nervous	1002
Disgust	disgusted appalled displeased fed up repulsed revolted scandalized sickened sick and tired turned off ew yuck	432

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 emotional vocabulary (min 200 characters)
This is a picture of a man in a purple suit. He is wearing blue sunglasses 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

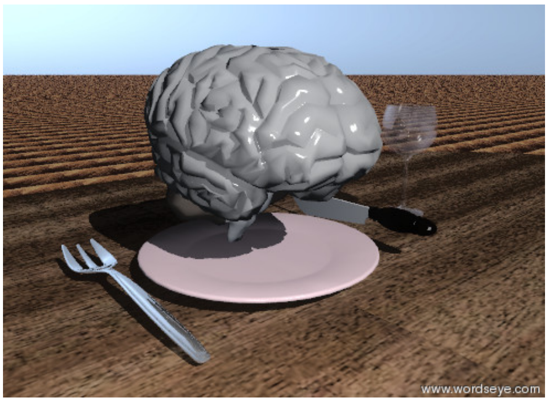
It also enables the rest of the users to comment on the picture and their impressions about it. Our goal is to be able to accurately set a mood for the scene based on this kind of data.

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 vocabulary (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 turkers 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.



Please write a literal description (3-5 sentences) that could work as a caption for the previous image. Avoid the use of emotional vocabulary (min 200 characters)
Right in the middle of a half sown wheat field is a human brain on a plate. The wheat is perfectly lined up on a perfect blue sky day. It is being prepared to be served up for dinner with an elegant glass of wine to accompany it.
Please choose the mood that relates the most to the displayed picture:
Disgust
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.
It is kind of gross to think about one eating an organ. Although it is really done around the world as a delicacy, it does not look that tempting at all. Especially the thought of eating a human brain.

Table 3: Real completed task

Please write a literal description (3-5 sentences) that could work as a caption for the previous image. Avoid the use of emotional vocabulary (min 200 characters)
Green, yellow and white lights surround a large black gate, an entrance to something dark and intriguing. There is a figure in front of the door that appears to be a child guarding entrance dearly.
Please choose the mood that relates the most to the displayed picture:
Fear
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.
I chose 'fear' to best describe this image because it is very eerie. It is a very unpleasant image.

Table 4: Real completed task

Class	# pictures	# 3/3 tags
Happiness	731	264
Sadness	267	24
Anger	69	9
Surprise	430	39
Fear	382	30
Disgust	101	9

Table 5: WordsEye examples per class

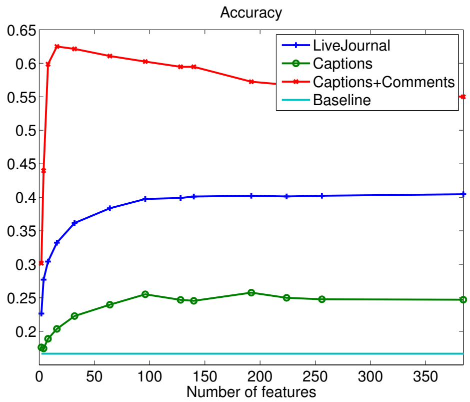


Figure 1: Accuracy of the 6-class classification problem over the WordsEye caption and comment dataset

(f) TF-IDF of emoticons. More than 50 distinct emoticons (clustered into 18 different groups according to a classification found in Wikipedia [6]) were detected in our dataset, and their TF-IDF scores were included as features.

5. METHODOLOGY

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 sets and, using the selected features, each pair of training and test set will be reduced to the feature vector lengths: 2, 4, 8, 16, 32, 64, 96, 128, 140, 192, 224, 256, 384 and 500. For each vector length, a search over the SVM cost parameter C is performed using 5-fold cross-validation. The C value that maximizes the accuracy of the model is chosen and using it the SVM is trained and tested.

6. FEATURE SELECTION

Given the high dimensionality of our prediction problem, it is necessary to apply feature selection techniques to build a smaller and more efficient model. The minimum redundancy maximum relevance (mRMR) technique [15] is used here for its effectiveness on selecting relevant features and discarding those redundant with the ones already selected. The method is executed over each 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 rates. The ranking is computed by the average min-redundancy max-relevance score of each feature over each partition and fold. From left to right the columns contain the rankings for the LiveJournal dataset, the WordsEye dataset using exclusively the captions and discarding the comments, and the full WordsEye dataset.

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 ninetieth 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 *Digusted* 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

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

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

	LiveJournal (128 feat.)			Captions (96 feat.)			Caption and Comments (16 feat.)		
Mood	Precision	Recall	F-score	Precision	Recall	F-score	Precision	Recall	F-score
Happy	0.46±0.02	0.55±0.02	0.50±0.02	0.26±0.08	0.26±0.1	0.25±0.09	0.59±0.04	0.81±0.06	0.67±0.04
Sad	0.49±0.02	0.49±0.02	0.49±0.02	0.21±0.09	0.22±0.07	0.21±0.07	0.73±0.04	0.77±0.05	0.74±0.04
Angry	0.38±0.02	0.37±0.04	0.37±0.03	0.33±0.04	0.36±0.05	0.34±0.05	0.62±0.05	0.65±0.05	0.61±0.03
Surprised	0.32±0.02	0.29±0.02	0.30±0.02	0.20±0.06	0.18±0.08	0.18±0.06	0.64±0.09	0.44±0.09	0.50±0.08
Scared	0.38±0.01	0.37±0.03	0.37±0.02	0.26±0.08	0.24±0.07	0.24±0.07	0.68±0.05	0.73±0.06	0.69±0.04
Disgusted	0.38±0.02	0.35±0.03	0.36±0.03	0.27±0.04	0.28±0.05	0.26±0.04	0.63±0.07	0.35±0.09	0.42±0.07
Accuracy	0.40 ± 0.01			0.26 ± 0.02			0.63 ± 0.02		

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

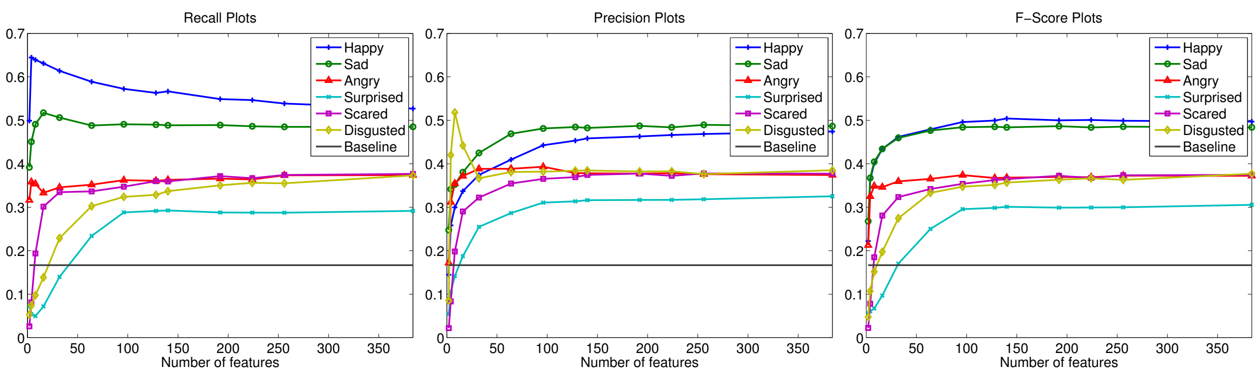


Figure 2: (Left) Recall, (Center) Precision and (Right) F-score of the 6-class classification problem over the LiveJournal dataset

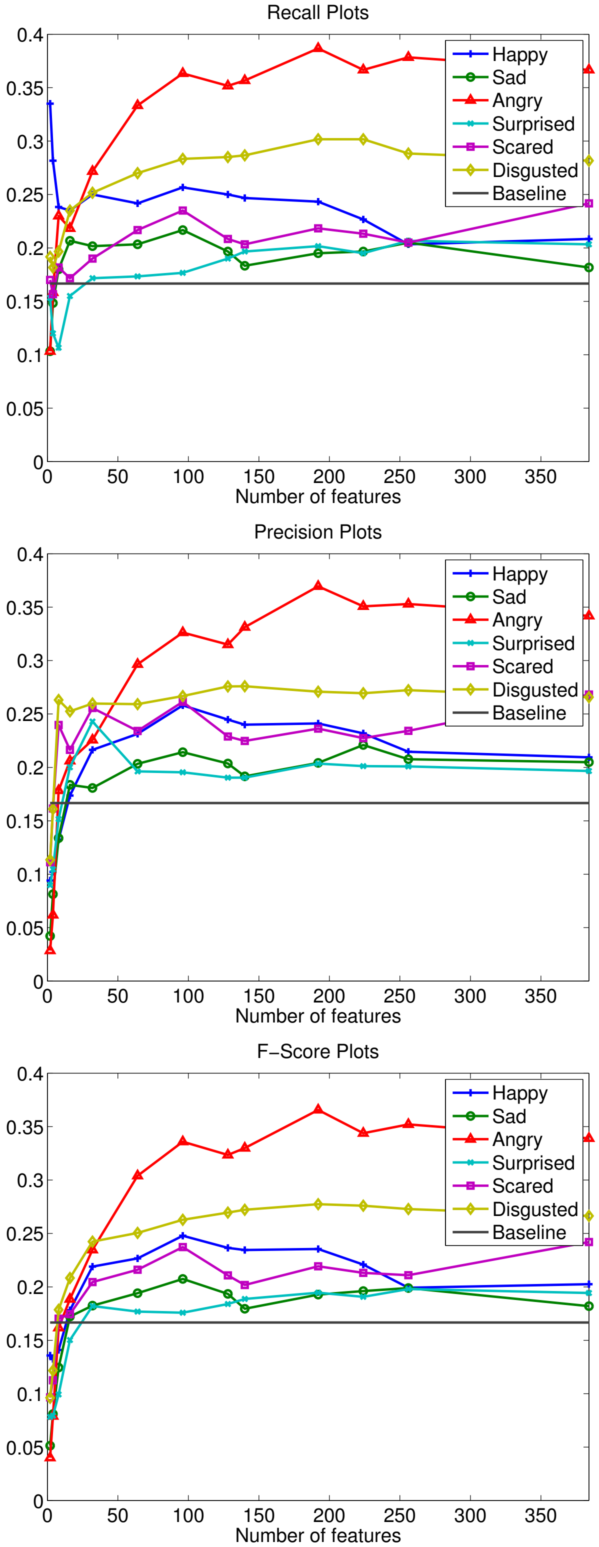


Figure 3: (Left) Recall, (Center) Precision and (Right) F-score of the 6-class classification problem over the WordsEye caption dataset

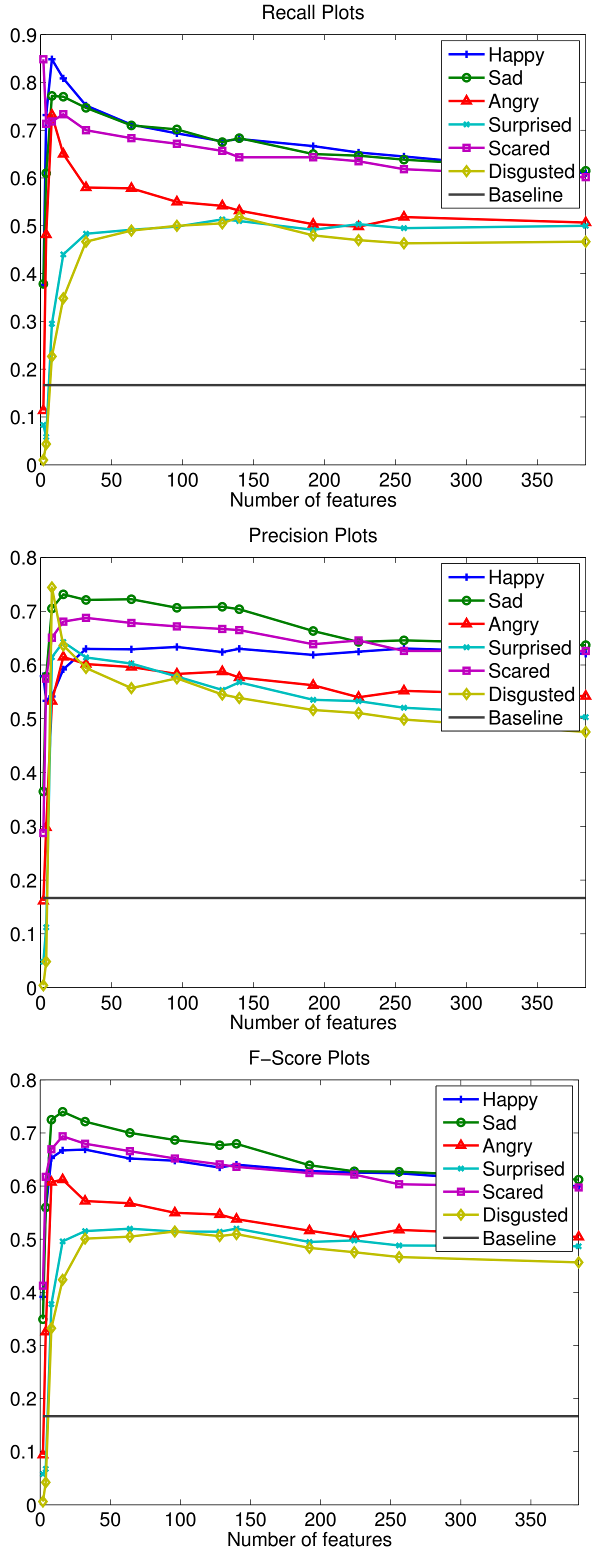


Figure 4: (Left) Recall, (Center) Precision and (Right) F-score of the 6-class classification problem over the WordsEye caption and comment dataset

	H	Sa	A	Su	F	D
Happiness	0.83	0.00	0.00	0.17	0.00	0.00
Sadness	0.33	0.67	0.00	0.00	0.00	0.00
Anger	0.17	0.00	0.83	0.00	0.00	0.00
Surprise	0.33	0.00	0.17	0.50	0.17	0.00
Fear	0.17	0.00	0.00	0.00	0.83	0.00
Disgust	0.00	0.17	0.00	0.33	0.17	0.33

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

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 data. We have tried directly testing a model trained on LiveJournal data on the image description data without promising results. We hope that employing more advanced methods for combining models will yield more success.

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

[1] R. Abbott, M. Walker, P. Anand, J. E. Fox Tree, R. Bowmani, and J. King. How can you say such things?!?: recognizing disagreement in informal political argument. In *Proceedings of the Workshop on Languages in Social Media*, LSM ’11, pages 2–11. ACL, 2011.

[2] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau. Sentiment analysis of twitter data. In *Proceedings of the Workshop on Languages in Social Media*, LSM ’11, pages 30–38. ACL, 2011.

[3] S. Baccianella, A. Esuli, and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*. European Language Resources Association (ELRA), may 2010.

[4] R. Balasubramanyan, W. W. Cohen, D. Pierce, and D. P. Redlawsk. What pushes their buttons?: predicting comment polarity from the content of political blog posts. In *Proceedings of the Workshop on Languages in Social Media*, LSM ’11, pages 12–19. ACL, 2011.

[5] R. A. Calix, S. A. Mallepudi, B. C. B. Chen, and G. M. Knapp. Emotion recognition in text for 3-d facial expression rendering, 2010.

[6] W. contributors. List of emoticons. In *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 2012.

[7] B. Coyne and R. Sproat. Wordseye: an automatic text-to-scene conversion system. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, SIGGRAPH ’01, pages 487–496, 2001.

[8] D. Das and S. Bandyopadhyay. Word to sentence level emotion tagging for bengali blogs. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, ACLShort ’09, pages 149–152, 2009.

[9] P. Ekman. Basic emotions. *The Handbook of cognition and emotion.*, pages 45–60, 1999.

[10] A. Esuli and F. Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC ’06)*, pages 417–422, 2006.

[11] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, 2008.

[12] C. Fellbaum, editor. *WordNet An Electronic Lexical Database*. The MIT Press, Cambridge, MA, May 1998.

[13] G. A. Miller. Wordnet: A lexical database for english. *Communications of the ACM*, 38:39–41, 1995.

[14] G. Mishne. Experiments with mood classification in blog posts. In *1st Workshop on Stylistic Analysis Of Text For Information Access*, 2005.

[15] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE transactions on PAMI*, 27(8):1226–1238, 2005.

[16] C. Strapparava and R. Mihalcea. Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied computing*, SAC ’08, pages 1556–1560. ACM, 2008.

[17] Y. R. Tausczik and J. W. Pennebaker. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54, Mar. 2010.

[18] C. Whissell. The Dictionary of Affect in Language. In R. Plutchik and H. Kellerman, editors, *Emotion: theory, research and experience*, volume The measurement of emotions. Academic Press, London, 1989.