Analyzing #emotions in Tweets

Naveen Kumar
IMS,University of Stuttgart,
Germany
kumarnn@ims.uni-stuttgart.de

Aswathy Velutharambath
IMS,University of Stuttgart,
Germany
veluthay@ims.uni-stuttgart.de

Abstract

Micro-blogging services like twitter are now-a-days being used by people to express their daily thoughts, feelings and emotions in real-time. Automatic detecting of emotions from such short texts provide powerful insights to the pattern of human communication over internet.In this project we are trying to investigate the pattern of emotions in twitter messages(tweets). We use a supervised classification method and experiment with a variety of features like unigrams, bigrams, NRC emotion dictionary, negation and discourse features.

1 Introduction

In recent years, social media is being used extensively by individuals for expressing their emotions. It is gaining lot of attention within NLP community as it provides large amount of data for opinion mining and emotion analysis. Twitter is one source of interest which allows only short text messages. Wide range of emotions are expressed through these tweets.

In this project we are focusing on emotion analysis on short text messages. We are approaching this emotion analysis task as a classification problem with different classes as Happy , Sad, Anger, Disgust, Fear and Surprise.

2 Method

2.1 Preprocessing

We use the data crawled from twitter as our training data. This labeled dataset is preprocessed to remove junk information and to obtain a better structured text. Following are the main rules applied as part of the pre-processing steps :

- Re-tweeted messages contain RT tags. We removed such tags so that identification and removal of duplicate tweets will be made easy.
- Usernames in tweets usually starts with @ symbol (e.g. @MyCoolSelf). All strings starting with @ are replaced by [USRNAME].
- Urls present as part of the tweets are also replaced in a similar manner by [URL].
- All hash-tags used for crawling are removed from the tweet in-order to avoid biasing for the hashtag class.
- After the above steps are completed, we removed the duplicate tweets by using cosine similarity.

  Original Tweet : *Today's Feature: "How Am I Doing? Fine, but Not Really" by Le Ann Trees #grief https://t.co/BN3eA3oghl https://t.co/1zfqMxDlpG*
  Preprocessed Tweet : *today's feature "how am i doing fine but not really" by le ann trees [URL]*

2.2 Perceptron

We use a supervised classification method by perceptron learning. Since we have to classify multiple emotion classes multi-class perceptron is used. We implemented multi-class perceptron using the idea of winning perceptron of one out of N perceptrons where N is number of classes. In training, perceptron algorithm is made to run K number of iterations. Each perceptron have weights for each feature vector where each weight is initialized as zero followed by updating them in each iteration depending upon the winning perceptron. In testing, winning perceptron refers to the predicted label using feature vector and corresponding weights for it.

Features Used	Anger	Sad	Happy	Surprise	Fear	Disgust	Micro F-Score	Macro F-Score
Ngram	0.45	0.57	0.79	0.30	0.59	0.23	0.66	0.49
Ngram + NRC	0.44	0.57	0.78	0.30	0.58	0.14	0.65	0.47
Ngram + NRC + Negation	0.44	0.57	0.78	0.32	0.59	0.18	0.65	0.48
Ngram + Negation + Discourse	0.52	0.64	0.82	0.36	0.62	0.26	0.71	0.54
Ngram + Negation + Discourse + Smiley	0.50	0.64	0.82	0.36	0.61	0.32	0.71	0.54

Table 1: Contributions of different features to the f-score of different classes, the micro and the macro F-scores

2.3 Feature sets

We implement different features for emotion classification which includes basic features like ngrams, advanced features like discourse connectives and also external lexical resources.

Ngrams. As a feature, we use unigrams and bigrams extracted from the tweet.

Emotion dictionary. The NRC Emotion Lexicon[1] consists of a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). The annotations were manually done by means of crowdsourcing.

We use the emotions as well as the two sentiments as features(eg.NRC-anger).

Negation feature. We use negation feature to identify the tweets which contain negated phrases like '*hardly surprised*' or '*not really happy*'. Such phrases can result in wrong classification if only unigrams and bigrams are considered. We add 'Negation' as a feature,if a negation word occurs in the tweet.

Discourse feature. We consider few discourse features like conditionals and connectives mentioned in Mukherjee and Bhattacharyya (2012). Presence of the conditional like *if* , *might* etc in the sentence has some bias towards the polarity of the sentence, hence we add it as a feature.

We also add the Conj_Fol as features where Conj_Fol is the set of conjunctions that give more importance to the discourse segment that follows them.(Mukherjee and Bhattacharyya ,2012).

Smiley feature. People generally resort to smileys to express emotions especially in short text messages. Here we use an external lexicon[2] of smileys in which which smileys are categorized as *Extremely Positive* , *Positive*, *Neutral*, *Negative* and *Extremely Negative*. We add the smiley and the category(eg :D-Extremely-Positive) as one feature

3 Experiment

3.1 Experimental setup

Data. We use the data crawled from twitter using different emotion classes as hastags and also hastags of related emotions (eg, #happy and # happiness). This data was used as the training data as they are already annotated with the hashtags.

Feature Selection. In order to select the best features, we add the features incrementally to the model and evaluate using micro and macro F-score. Table 1 shows the result of this feature selection experiment.

From the table it can be seen that after adding the NRC feature, the micro and macro F-score reduced and in the next step, even on adding Negation feature there was no considerable improvement in the F-scores. So we removed the NRC feature and added Discourse features, which improved the micro and macro F-scores significantly.

In the case of Smiley feature, even though it did

[1] http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

[2] https://github.com/mayank93/Twitter-Sentiment-Analysis/blob/master/phrase-level-analysis/code/emoticonsWithPolarity.txt

Rank coverage/ score	**Anger**	**Sad**	**Happy**	**Surprise**	**Fear**	**Disgust**	**Micro F-Score**	**Macro F-Score**
1	0.49	0.63	0.81	0.36	0.61	0.32	0.71	0.54
2	0.78	0.84	0.88	0.81	0.81	0.87	0.84	0.83
3	0.88	0.92	0.93	0.9	0.9	0.94	0.91	0.91
4	0.94	0.96	0.96	0.95	0.95	0.96	0.96	0.96
5	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.98
5	1	1	1	1	1	1	1	1

Table 2: Rank coverage and score per category

not contribute in improving the micro and macro F-scores, the F-score of the class *disgust* is increased significantly which was otherwise very poor. So we decided our final feature set to be *Ngram* + *Negation* + *Discourse* + *Smiley*.

3.2 Error Analysis

The results shown in Table 1 clearly indicates poor performance for the classes *Disgust* and *Surprise*. It could be due to data sparsity as the classes are not frequent in training as well as test data.

On manually analyzing the results, we found that multiple emotions are expressed in the same tweet. For example, from the confusion matrix given in Table 3, we can see that the class *Disgust* is predicted as *Sad* 9 times. Also the number of times *Surprise* is predicted as *Happy* is more than *Surprise* itself.This could be because both the emotions are present in the same tweet. In the next section,we investigate further with a second experiment on the data by considering more than one top ranked emotions.

Gold / Predicted	Disgust	Sad	Anger	Happy	Fear	Surprise
Disgust	4	9	0	3	1	1
Sad	2	821	55	184	84	13
Anger	0	133	222	96	60	4
Happy	1	282	48	2321	142	36
Fear	0	137	44	155	519	8
Surprise	0	43	8	78	22	61

Table 3: Confusion matrix result for test data

3.3 Rank Coverage

Until now we predicted the emotion class which was scored highest by the classifier and calculated performance based on the predicted class. However, if there are multiple closely related emotions expressed in the same tweet, the tweet could be classified as the related class and not as the gold class. To investigate on this aspect, we modified the algorithm to consider the top n predicted classes and update the confusion matrix.

As explained in Algorithm 1,for example if the gold class is present in the top 2 predicted classes, we increment the true positive count for both predictions by 0.5 . In the case where gold label is not present in the top predictions, we increment the false negative for the gold class by 1 and increment the false positive for the predictions by the fraction.

Algorithm 1 Evaluation in Ranking Scenario

$PL \leftarrow$ *List of top N predicted labels*
$G \leftarrow$ *Gold label for the data*
if PL $contains$ G **then**
  Increment TP for all PL by (1/(Size of PL))
else
  Increment FN for G by 1
  Increment FP for all PL by (1/(Size of PL))
end if

The results displayed in Table 2 shows a significant improvement in scores when the first two categories are considered. Especially for the categories like disgust and surprise for which data is relatively sparse the results improved drastically. This could imply a strong co-occurrence relation between emotions in the tweets. Further investigation in this direction could be an interesting focus for research.

4 Summary

In this paper, we did emotion analysis on Tweets using a supervised classification method. On the results obtained after feature engineering we did an error analysis to investigate further on false positives. We used a rank coverage method, in which we considered top *N* classes, check for the gold class and update confusion matrix as per the new algorithm. It was observed that the result improved drastically when the first two ranks where considered which suggested a possibility of co-occurrence of emotions in tweets.

5 Future scope

For further improvement of the result, we would like to concentrate on the feature improvements by experimenting with features including filtering out the features based on mutual information. Another possibility is to use standard polarity dictionaries to segregate emotions in the feature space.

The rank coverage method and its implications towards the co-occurrence of emotions could be a focus for further research

References

Subhabrata Mukherjee and Pushpak Bhattacharyya. 2012. *Sentiment Analysis in Twitter with Lightweight Discourse Analysis*. Proceedings of COLING 2012: Technical Papers, pages 1847–1864