Team LAB nlp. Twitter mining

# Progress report

**Emotion Analysis:** Looking into ways people express their emotion using text

Seed words for each emotion category e.g. happy, enjoy, pleased are seed words for happy. Afraid, scared, panic for fear category.

|  |
| --- |
| I have to look at life in her perspective, and it would break anyone’s heart. (*sadness, high*) |
| We stayed in a tiny mountain village called Droushia, and these people brought hospitality to incredible new heights. (*surprise, medium*) |
| But the rest of it came across as a really angry, drunken rant. (*anger, high*) |

Annotation via 4 judges in order to avoid biasness which annotated the text using sample sentences

Ekman 6 basic emotions other than mixed and no emotion

I felt bored and wanted to leave at intermission, but my wife was really enjoying it, so we stayed (mixed emotion)

Identify features that distinctly categorize emotional sentences in text and are not likely to be found in non-emotional sentences. (Emotion words). For recognition of emotion words lexical resources of General inquirer/WordNet were used.

|  |  |  |
| --- | --- | --- |
| GI Features | WN-Affect Features | Other Features |
| Emotion words Positive words Negative words Interjection words Pleasure words Pain words | Happiness words Sadness words Anger words Disgust words Surprise words Fear words | Emoticons Exclamation (“!”) and question (“?”) marks |

**Emotion extraction in Twitter:**

Given the dataset as dev (gold) standard of individual records comprising emotion, language, date & time, tweet-id, tweet, latitude-longitude, state, country**.**Dataset dev-predicted (predicted) contains emotions that are presumably classified by model, against every tweet in dev (gold) dataset in single line.  
Analyzing the corpus, we observed 8 different emotion in dev.csv (gold) corpus with distribution of:  
**cat dev.csv | awk '$1 {print $1}' | sort | uniq -c**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Emotions** | Happy | Sad | Anger | Fear | Surprise | Disgust | Trust | Love |
| **Count** | 179484 | 76380 | 42960 | 31078 | 13102 | 1190 | 757 | 66128 |

For dev-predicted

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Emotions** | Happy | Sad | Anger | Fear | Surprise | Disgust | Trust | Love |
| **Count** | 329299 | 53170 | 10241 | 14387 | 1939 | 533 | 1050 | 460 |

For evaluation task we calculated precision, recall, F-score at micro and macro level for all emotions. For micro level, we calculated precision, recall and F-score for each emotion individually.   
For instance micro precision of happy is calculated as**:**

**Precision ­happy**­=

**Recall ­happy**­=

**­F-Score ­happy ­= 2**

At macro level we summed up all individual scores of micro level for each emotion and divided by total number of emotions that we considered.

**Precision­ (macro) =**

**Recall (macro) ­=**

**F-Score ­ (macro) = ­= 2**

**Feature extraction:**

Trimming vocabulary: removing stop words, symbols, hashtags, Retweets

Stemming: (Reduction of inflectional forms to base forms)

Class definition according to em0tion categories  
 (8 classes for emotion, 1 for no emotion and 1 for neutral)