5.6 Creativity

Since this part requires more ideas on deep-learning module for analysis of comments. Our group has decided using Emoticon Annotated to make categories of the corpus from Tweet and Reddit.

In our instance, we present a framework of lexicon, which contains disambiguated entity with a three dimensional probability distribution as “negative”, “positive”, and “neutral” polarities. At the beginning, we crawl tweet comments and store each as a text file (\*.txt format). We have got around 6700 records. Then we develop a system with automatically detection what people feel about from records:

1. Percentage of people who feel happy of their candidates selection
2. Percentage of people who feel annoy of their candidates selection
3. Percentage of people who are not care about the selection

Classifying text message based on their emotion and sentiment comment.

After collecting labeled tweets, we proceed Date Pre-processing.

First, we define our “Emoticon” features. Below is the full list of emoticons that we used in our system (Table 1):

|  |  |
| --- | --- |
| Category | Emoticons |
| Happy Emoticons | ':)', ';)', '=)', ':]', ':P', ':-P', ';P', ':D', ';D', ':>', ':3', ':-)', ';-)', ':^)', ':o)', ':~)', ';^)', ';o)', ":')", ':-D', ':->' |
| Sad Emoticons | ':(', '=(', ':-(', ':^(', ':o(', ';^(', ":'(", ':-<' |

**Table 1**

Based on the emotional release, we got the result (Table 2):

|  |  |
| --- | --- |
| Negative Attitude | 1201 |
| Positive Attitude | 1419 |
| Neutral Attitude | 4115 |
| Total | 6735 |

**Table 2**

The Weka workbench is an organized collection of state-of-the-art machine learning algorithms and data preprocessing tools.

The result in table2 has been converted to Weka format and we later use “Vote” classifier to test cross validation of our result.

Finally, the fluctuant (both positive and negative) rate takes 38.9% possession.

In order to make a comparison, we have applied Sentiment Analysis as well.

Here is the sentiment list (Table 3):

|  |  |
| --- | --- |
| Category | Text |
| Happy Text | 'elated', 'overjoyed', 'enjoy', 'excited', 'proud', 'joyful', 'happy', 'blessed', 'amazing', 'wonderful', 'excelent', 'delighted', 'enthusiastic', 'calm', 'peace', 'silent', 'serene', 'consent', 'convince', 'satisfied', 'relax' |
| Sad Text | 'nervous', 'anxious', 'tension', 'afraid', 'fearful', 'distress', 'stress', 'angry', 'annoy', 'tense', 'bother', 'disturb', 'irritate', 'mad', 'furious', 'sad', 'sorrow', 'hapless', 'fatigue', 'gloomy', 'miserable', 'hopeless', 'depress', 'fatigued', 'unhappy', 'laugh', 'not' |

**Table 3**

And corresponding result list (Table 4):

|  |  |
| --- | --- |
| Negative Attitude | 748 |
| Positive Attitude | 158 |
| Neutral Attitude | 5829 |
| Total | 6735 |

**Table 4**

As same process, the fluctuant (both positive and negative) rate takes 13.5% possession in sentiment text analysis.

By competition in “Emoticons” and “Text”, system will have higher sentiment detection when it implements “Emoticons” features (38.9%) rather than “Text” features (13.5%).

This is our creative idea for sentiment analysis.

Since the dataset we crawled is a sample (little size). We may enhance our emoticon detection to achieve further higher accuracy of data analysis.

Thank you