

TM-A3

Topic:Sentiment Analysis

Group 33: Wei Chen and Jialiang Wang

December 2021

1 Introduction

Sentiment Analysis is the process of analyzing, processing and extracting subjective text with emotion by using natural language processing and text mining techniques. In this assignment, the task is using sentiment analysis technique to judge a tweet whether is positive, negative or neutral.

2 Data description

Each data set has the same structures, with a string of numeric labels, the sentiment which consists of three criteria: neutral, positive, and negative (three-point scale). Also, the final columns is the tweets. For instance, the data set *Spending my Saturday getting my car serviced is definitely the most enjoyable thing I could do with my time.* is negative in sentiment analysis. *It's Super Bowl Sunday, pastors. Get your Jesus Jukes ready! Guilt is an awesome motivator!* is positive, and *day God known Sunday week Question and Answer <http://t.co/tz18q9Brvy> prayer church* is neutral.

But there are some duplicated items during our data pre-processing. In the datasets 2016, the 2016-testA has 51 duplicated items and 2016-trainA has 19 duplicated items. In the datasets 2015, the 2015-testA has 7 duplicated items. In the datasets 2013, the 2013-devA has 4 duplicated items, the 2013-testA has 2 duplicated items and the 2013-trainA has 29 duplicated items.

| | 2016-testA | 2016-trainA | 2016-devtestA | 2016-devA |
|----------|------------|-------------|---------------|-----------|
| Neutral | 10314 | 2036 | 681 | 765 |
| Positive | 7046 | 3082 | 994 | 843 |
| Negative | 3221 | 863 | 325 | 391 |
| Sum | 20581 | 5981 | 2000 | 1999 |

| | 2015-trainA | 2015-testA |
|----------|--------------------|-------------------|
| Neutral | 253 | 985 |
| Positive | 170 | 1034 |
| Negative | 66 | 364 |
| Sum | 489 | 2383 |

| | 2014-sarcasmA | 2014-testA |
|----------|----------------------|-------------------|
| Neutral | 13 | 669 |
| Positive | 33 | 982 |
| Negative | 40 | 202 |
| Sum | 86 | 1853 |

| | 2013-devA | 2013-testA | 2013-trainA |
|----------|------------------|-------------------|--------------------|
| Neutral | 737 | 1513 | 4568 |
| Positive | 573 | 1473 | 3633 |
| Negative | 340 | 559 | 1454 |
| Sum | 1650 | 3545 | 9655 |