ARI5121
Applied NLP
Emotion Detection in Twitter Data

Gabriel Sammut
gabriel.sammut.13@um.edu.mt
Academic Year 2018/2019

1 Introduction

This report covers respective documentation concerning ARI5121, and is split into two separate components:

1. Emotion Detection in Twitter Data using Naive Bayes classifiers.
2. Using Conditional Random Fields for POS and NER tagging.

Both subsets have been evaluated separately from each other, using different datasets.

2 Datasets

Each of the above two experiments was conducted on separate datasets, as follows:

1. Twitter data, containing user tweets. Each tweet is flagged with a number of emotions, ranging from one to eight [1]. Each tweet can pertain to a number of the following emotions:
    a. Anger
    b. Anticipation
    c. Disgust
    d. Fear
    e. Joy
    f. Sadness
    g. Surprise
    h. Trust
2. Newspaper text snippets, pre-annotated with respective tags. A subset of the data is annotated with Part of Speech tags, another annotated with name entity recognized tags.

3 Experiment 1

The first experiment deals with emotion identification of Twitter messages. A total of eight Naive Bayes Classifiers are established, each dedicated and trained on a particular emotion. The experiment itself was split into three groupings, as detailed below. Achieved results were evaluated on a set of reserved test data, not used in the training of models.

The first part of the experiment evaluates each emotion classifier on an individual basis - each trained model is evaluated individually, irrespective of other emotions. The following metric scoring was achieved for each emotion classifier:

Emotion	Accuracy	Precision	Recall	FScore
Anger	70.3%	75.1%	79.9%	77.4%
Anticipation	59.6%	68.0%	65.0%	66.5%
Disgust	64.8%	61.0%	67.7%	64.2%
Fear	62.5%	54.4%	51.6%	53.0%
Joy	64.5%	53.5%	58.7%	56.0%
Sadness	60.6%	62.3%	69.3%	65.6%
Surprise	64.9%	32.9%	28.8%	30.7%
Trust	68.7%	55.7%	49.7%	52.5%

Table 1 - Individual Emotion Evaluation

The second part of the experiment attempts to combine two emotion classifiers together, in an attempt to improve a particular emotion's scoring metric. By comparing two Naive Bayes models together, an attempt is made to boost individual model scoring. Pairwise assignment is made as follows:

- e1 | e2
- e1 | !e2
- !e1 | e2
- !e1 | !e2

The first emotion to be evaluated in a pairwise fashion is the surprise emotion, evaluated as follows:

Pairwise Emotion	Accuracy	Precision	Recall	FScore
Anger	57.8%	31.9%	49.9%	38.9%

Anticipati on	57.8%	31.9%	49.9%	38.9%
Disgust	57.8%	31.9%	49.9%	38.9%
Fear	57.8%	31.9%	49.9%	38.9%
Joy	57.8%	31.9%	49.9%	38.9%
Sadness	57.8%	31.9%	49.9%	38.9%
Trust	57.8%	31.9%	49.9%	38.9%

Table 2 - Surprise Pairwise Emotion Evaluation

Pairwise classification for the 'Surprise' label seems to have a decrease on evaluated accuracy, whilst achieving a notable increase in achieved Recall and F-Score metrics. A second emotion, particularly 'Disgust is also tested out:

Pairwise Emotion	**Accuracy**	**Precisi on**	**Recall**	**FScore**
Anger	65.2%	60.9%	70.8%	65.5%
Anticipati on	65.2%	60.9%	70.8%	65.5%
Fear	65.2%	60.9%	70.8%	65.5%
Joy	65.2%	60.9%	70.8%	65.5%
Sadness	65.2%	60.9%	70.8%	65.5%
Surprise	65.2%	60.9%	70.8%	65.5%
Trust	65.2%	60.9%	70.8%	65.5%

Table 3 - Disgust Pairwise Emotion Evaluation

The 'Disgust' emotion exhibits an improvement in all evaluated metrics by comparing it with another emotion.

The third experiment milestone attempts to identify the best pairwise emotion combination of all possible eight emotions. A brute force mechanism was attempted, by comparing all possible pairwise combinations (56 possible combinations). The best scored metrics are recorded were achieved for the emotion 'Anger', given 'Anticipation'':

Accuracy	**Precision**	**Recall**	**FScore**
70.6%	77.0%	76.7%	76.8%

Table 4 - Best Pairwise Combination

4 Experiment 2

The second experiment dealt with part of speech tagging and name entity recognition on two separate, distinct datasets. In particular, CRF (conditional random fields) models were trained for both cases. Each training dataset had a respective test data set, used to eventually evaluate the trained models

For both trained models (POS and NER), the following word features were loaded:

1. Word in lowercase
2. Last 3 characters of word
3. Last 2 characters of word
4. Boolean flag if word is upper cased
5. Boolean flag if word is title cased
6. Boolean flag if word is a digit
7. Actual training (POS)/(NER) tag
8. Word before in lowercase
9. Boolean flag if word before is title cased
10. Boolean flag if word before is upper cased
11. Boolean flag if word before is a digit
12. Word after in lowercase
13. Boolean flag if word after is title cased
14. Boolean flag if word after is upper cased
15. Boolean flag if word after is a digit

Both models were trained using the following parameter configurations:

Parameter	**Value**
c1	0.1
c2	0.01
max_iterations	50
possible _transitions	True

Table 5 - CRF Hyper Parameter Configuration

Both CRF trained models exhibited exceptional performance once evaluated on the test data sets, as shown below:

Model	**Accuracy**	**Precision**	**Recall**	**Fscore**
POS	100%	100%	100%	100%
NER	100%	100%	100%	100%

Table 6 - CRF Metric Scoring

For both CRF models, a perfect scoring for both trained models was achieved. Upon further tweaking of each individual model hyper-parameters, it was discovered that the most important features were those concerning the actual word itself. Features concerning the word before and after the actual word seemed to show little effect on achieved F-Score. These extra features could decrease overall model training time, if removed.

5 Conclusion

In this report, two different Natural Language Processing techniques were implemented on two separate scenarios, ranging from twitter tweet emotion prediction using Naive Bayes Models, as well as part of speech tagging and named entity recognition through usage of conditional random fields. Both models were evaluated under Recall, Precision and F-Score metrics, exhibiting varied, yet significant performance.

6 References

[1]Romanklinger.de, 2018. [Online]. Available: http://www.romanklinger.de/ssec/ssec-aggregated-withtext.zip. [Accessed: 16- Jun- 2018].

[2]"eldrad294/ARI5092_Practical_Assignment_75_1", GitHub, 2018. [Online]. Available: https://github.com/eldrad294/ARI5092_Practical_Assignment_75_1/tree/master/data/pos. [Accessed: 16- Jun- 2018].