Assignment two: Detection opinion spam

Mike Vink

October 27, 2020

Contents

1 Introduction

2 Dataset construction

2.1 Selection of data

The dataset used in this paper is derived from the publicly available dataset of *negative* deceptive opinion spam contributed by [1], containing 400 gold standard deceptive negative reviews of 20 popular Chicago hotels, and 400 truthfully negative reviews. This data consists of text files that were used to construct our dataset, no cleaning of the data was necessary.

2.2 Preprocessing of the review text data

2.3 Construction of derived features

*

3 Deception classification results

Table 1: Classifier performance on the training set (640 reviews Cross. Val.) and the test set (160 reviews Held Out)

Model	Features	Accuracy	P	R	F
Multinomial naive bayes	Unigrams				
	Bigrams				
Logistic regression	Unigrams				
	Bigrams				
Random forests	Unigrams				
	Bigrams				

3.1 Multinomial naive bayes

3.2 Logistic regression

3.3 Classification trees and random forests


```
{'rs_cv': {'mean_fit_time': array([0.3270334 , 0.32086271, 0.33634794, 0.33707774, 0.
       0.4065218 , 0.36038804, 0.3671118 , 0.3812362 , 0.37610584]), 'std_fit_time':
       0.01974343, 0.05183415, 0.01868734, 0.06508651, 0.02701626]), 'mean_score_time
       0.00322396, 0.00316936, 0.0030092 , 0.00395298, 0.0031541 ]), 'std_score_time'
       1.33357797e-04, 2.85231207e-04, 3.18291633e-04, 9.28722997e-05,
       9.68385465e-04, 1.89983660e-04]), 'param_ccp_alpha': masked_array(data=[0.0035
                  0.003059188892211335, 0.008156661396612185,
                  0.0033752376170207024, 0.008655144702494299,
                  0.005368112045171204, 0.002797735630258966,
                  0.006744128519064042, 0.00582901885932175],
            mask=[False, False, False, False, False, False, False, False,
                  False, False],
       fill_value='?',
            dtype=object), 'params': [{'ccp_alpha': 0.0035087283040104557}, {'ccp_alp
       0.68125, 0.68125, 0.6875 ]), 'split1_test_score': array([0.675  , 0.7    , 0.6
```

0.66875, 0.7125 , 0.7125]), 'split2_test_score': array([0.575 , 0.65   , 0.6
0.59375, 0.6875 , 0.61875]), 'split3_test_score': array([0.5625 , 0.5375 , 0.5
0.55625, 0.575 , 0.55   ]), 'mean_test_score': array([0.61875 , 0.6421875, 0
0.6328125, 0.625   , 0.6640625, 0.6421875]), 'std_test_score': array([0.05038
0.06240227, 0.0670784 , 0.0519164 , 0.05273293, 0.06333428]), 'rank_test_score
[66, 250]]), 'confusion_test': array([[46, 21],
[34, 59]])}

3.4 Model comparisons

4 Conclusions

5 References

References

[1] Myle Ott, Claire Cardie, and Jeffrey T Hancock. Negative deceptive opinion spam. In *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies*, pages 497–501, 2013.