

Predicting Opinions

Harshit Patel
2015MCS2338

June 2016

1 Introduction

The document contains results of various experiments conducted to predict opinions of politicians. The experiments are performed over **2280** tuples. **5-fold cross-validation** is used to determine accuracy of the models. Models used for experiments are **Decision Tree** and **Naive Bayes**. Results of **SVM** are not included in this report due to its very poor performance on data. Features were extracted from corpora through various methods like **Latent Dirichlet Allocation**, **set of n-grams with tf-idf** and **phrase2vec**.

2 Results

2.1 For features extracted using Latent Dirichlet Allocation

	Decision Tree	Naive Bayes
Accuracy 1	0.728	0.531
Accuracy 2	0.687	0.563
Accuracy 3	0.717	0.540
Accuracy 4	0.746	0.582
Accuracy 5	0.575	0.556
Average	0.690	0.554

2.2 For features extracted using set of n-grams with tf-idf

2.2.1 1-grams

	Decision Tree	Naive Bayes
Accuracy 1	0.627	0.544
Accuracy 2	0.620	0.542
Accuracy 3	0.698	0.542
Accuracy 4	0.700	0.544
Accuracy 5	0.672	0.547
Average	0.673	0.544

2.2.2 2-grams

	Decision Tree	Naive Bayes
Accuracy 1	0.680	0.544
Accuracy 2	0.717	0.577
Accuracy 3	0.691	0.531
Accuracy 4	0.691	0.520
Accuracy 5	0.678	0.503
Average	0.691	0.535

2.2.3 3-grams

	Decision Tree	Naive Bayes
Accuracy 1	0.733	0.565
Accuracy 2	0.697	0.517
Accuracy 3	0.715	0.521
Accuracy 4	0.802	0.553
Accuracy 5	0.734	0.561
Average	0.736	0.543

2.2.4 (1,2)-grams

	Decision Tree	Naive Bayes
Accuracy 1	0.668	0.544
Accuracy 2	0.724	0.549
Accuracy 3	0.698	0.551
Accuracy 4	0.714	0.548
Accuracy 5	0.609	0.545
Average	0.682	0.547

2.2.5 (1,2,3)-grams

	Decision Tree	Naive Bayes
Accuracy 1	0.763	0.544
Accuracy 2	0.742	0.542
Accuracy 3	0.678	0.542
Accuracy 4	0.730	0.544
Accuracy 5	0.736	0.547
Average	0.729	0.543

2.3 For features extracted using phrase2vec

	Decision Tree	Naive Bayes
Accuracy 1	0.659	0.542
Accuracy 2	0.733	0.521
Accuracy 3	0.659	0.544
Accuracy 4	0.675	0.573
Accuracy 5	0.639	0.508
Average	0.673	0.537

3 Observation

From the results, it can be seen that features built using **set of 3-grams with tf-idf** along with **Decision Tree** model gives the best accuracy of **0.802** and best average accuracy of **0.736**.

4 Future work

Improving model parameters, features and experiments using label propagation method.