Sentiment Analysis for Movie Reviews

Bharadwaj Tanikella, Rohit Pathak

- Binary classification (POS, NEG) on the basis of modified BOW and other RST features
- Dataset: Stanford AI (Large Movie Review Dataset)
  - 50k reviews (25k train and test each)
  - We chose 100 (random) training (50 neg and pos each) and 40 test
- Naïve Bayes Classifier (will try Perceptron)
- Features:
  - Nucleus BOW
    - nucleus = first `SpanNode` with nucleus prop in `nodelist.reverse`
  - Number of Relations (hypothesis: more 'evidence' is positive)
  - Relation BOW `{ ([relation], word): count }`
    - In combination with a relation, a word can weigh more or less
  - Relation POS `{ ([relation], POS): count }`
    - More adjectives in a relation may indicate positive review
  - 'Polarity' of Nucleus `{ ([POS/NEG]): count }`
- Evaluation Metrics:
  - Accuracy
  - F – measure
  - ROC curve & Cross Validation

Preliminary Results

Feature	Naïve Bayes		Perceptron	
	Accuracy	F-measure	Accuracy	F-measure
Bag of Words	0.50	0.000	0.50	0.000
Nucleus BOW	0.50	0.231	0.48	0.091
Num Relations	0.55	0.625	0.50	0.000
Relation BOW	0.60	0.556	0.53	0.174
Relation POS	**0.65**	**0.696**	0.50	0.000
Nucleus Polarity	0.53	0.536	0.50	0.091
…				
RST Combined	0.63	0.545	0.50	0.000

Disclaimer: The Perceptron maybe producing incorrect results.