Sentiment analysis for english text
2 Mart

Features: unigrams and bigrams for words and chars; count of polar words; count of emoticons

Classifiers: NB or SVM

Datasets: 1) 2K movie reviews 2) 7K tweets
Annotated for polarity.

Classes: Positive, negative.

Textual unit: document

Experiments:

Dataset	Features	Classifier	Accuracy
2K reviews	Word unigrams + word bigrams + char bigrams	Naive Bayes	0.82
7K tweets	Word unigrams + word bigrams + char bigrams	Naive Bayes	0.75
2K tweets + 7K reviews	Word unigrams + word bigrams + char bigrams	Naive Bayes	0.75

dataset	features					classifier	accuracy	duration (sec)
	Word ngram	Char ngram	Emoticon count	Polar word count	Polyglot polarity			
7K tweets	Uni + bigram	bigram	0	0	0	NB	0.75	224

7K tweets	Uni + bigram	bigram	0	0	0	SVM	0.75	227
7K tweets	Uni + bigram	bigram	1	1	1	SVM	0.76	
7K tweets	Uni + bigram	bigram	1	1	0	SVM	0.77	
7K tweets	Uni + bigram	bigram	1	0	0	SVM	0.75	
7K tweets	Uni + bigram	bigram	0	1	0	SVM	0.77	243
10K reviews	Uni + bigram	bigram	0	1	0	SVM	0.75	400
10K reviews	Uni + bigram	bigram	1	1	1	SVM	0.73	383
10K reviews	Uni + bigram	bigram	0	1	1	SVM	0.73	393
10K reviews	Uni + bigram	bigram	1	1	0	SVM	0.75	
10K reviews	Uni + bigram	bigram	0	0	0	SVM	0.74	