Sentiment Analysis

Sentiment Analysis

this movie was `great!` *would* `watch again` `+`

the movie was `gross` *and* `overwrought`, *but I* `liked` *it* `+`

this movie was `not` *really very* `enjoyable` `—`

▸ Bag-of-words doesn't seem sufficient (discourse structure, negation)

▸ There are some ways around this: extract bigram feature for "*not* X" for all X following the *not*

Pang et al. (2002)

	Features	# of features	frequency or presence?	NB	ME	SVM
(1)	unigrams	16165	freq.	**78.7**	N/A	72.8
(2)	unigrams	"	pres.	81.0	80.4	**82.9**
(3)	unigrams+bigrams	32330	pres.	80.6	80.8	**82.7**
(4)	bigrams	16165	pres.	77.3	**77.4**	77.1
(5)	unigrams+POS	16695	pres.	81.5	80.4	**81.9**
(6)	adjectives	2633	pres.	77.0	**77.7**	75.1
(7)	top 2633 unigrams	2633	pres.	80.3	81.0	**81.4**
(8)	unigrams+position	22430	pres.	81.0	80.1	**81.6**

▸ Simple feature sets can do pretty well!

▸ Learning alg. doesn't matter too much

▸ ME = "Maximum Entropy" = what we call Logistic Regression

Bo Pang, Lillian Lee, Shivakumar Vaithyanathan (2002)

Wang and Manning (2012)

▸ 10 years later — revisited basic BoW classifiers vs. other methods

Method	RT-s	MPQA
MNB-uni	77.9	85.3
MNB-bi	**79.0**	86.3
SVM-uni	76.2	86.1
SVM-bi	77.7	**86.7**
NBSVM-uni	**78.1**	85.3
NBSVM-bi	**79.4**	86.3
RAE	76.8	85.7
RAE-pretrain	77.7	**86.4**
Voting-w/Rev.	63.1	81.7
Rule	62.9	81.8
BoF-noDic.	75.7	81.8
BoF-w/Rev.	76.4	84.1
Tree-CRF	77.3	86.1

Before neural nets had taken off — results weren't that great

Kim (2014) CNNs `81.5` `89.5`

Wang and Manning (2012)