Sentiment Analysis

Sentiment Analysis



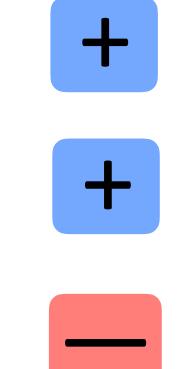
this movie was great! would watch again

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*



- the movie was gross and overwrought, but I liked it



Pang et al. (2002)



	Features	# of	frequency or	NB	ME	SVM
		features	presence?			
(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

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

- Simple feature sets can do pretty well!
- Learning alg. doesn't matter too much

Bo Pang, Lillian Lee, Shivakumar Vaithyanathan (2002)

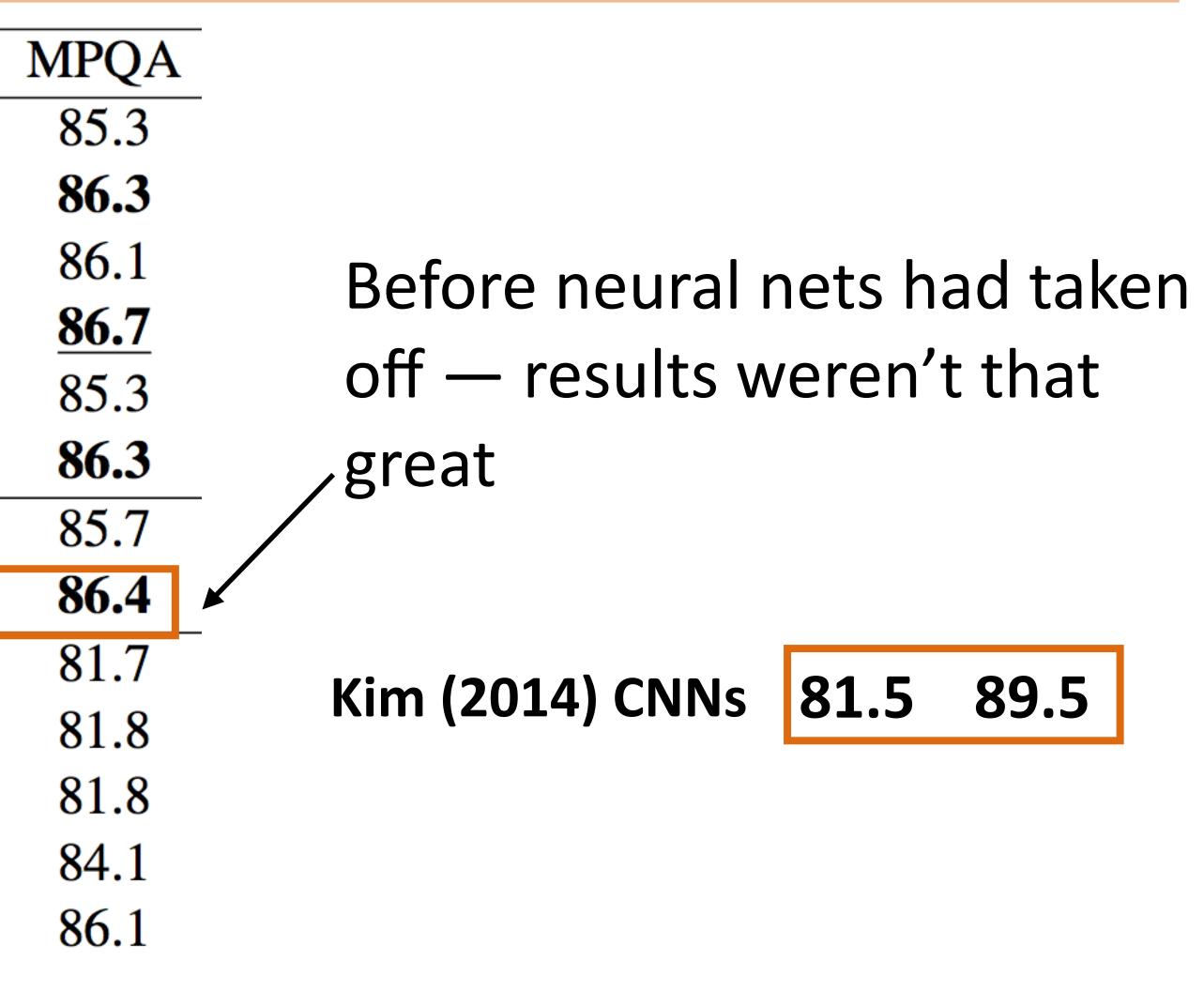






Wang and Manning (2012)

Method	RT-s	
MNB-uni	77.9	
MNB-bi	79.0	
SVM-uni	76.2	
SVM-bi	77.7	
NBSVM-uni	78.1	
NBSVM-bi	<u>79.4</u>	
RAE	76.8	
RAE-pretrain	77.7	
Voting-w/Rev.	63.1	
Rule	62.9	
BoF-noDic.	75.7	
BoF-w/Rev.	76.4	
Tree-CRF	77.3	
	MNB-uni MNB-bi SVM-uni SVM-bi NBSVM-uni NBSVM-bi RAE RAE-pretrain Voting-w/Rev. Rule BoF-noDic. BoF-w/Rev.	



Wang and Manning (2012)