Sentiment Analysis

Jaime Arguello
INLS 613: Text Data Mining
jarguell@email.unc.edu

October 25, 2015

Midterm

- Average: 87.07
- Median: 89.00

Definitions

- Subjectivity analysis: detecting whether a span of text describes the author's internal state (e.g., opinions, evaluations, emotions, speculations)
- Opinion mining: detecting whether a span of text expresses a positive/negative judgement
- Affect Detection: detecting whether a span of text conveys a particular emotion (e.g., anger, hope, disgust)

Applications

- Review summarization
- Recommendation systems
- Detecting “flames” in social media
- Summarization of multiple viewpoints
- Text-based forecasting or “now-casting”
- eRulemaking
- …

Challenges

Challenges

- Sarcasm
- Negation
- Modal verbs (e.g., could, should, would)
- Absence of “opinionated” text (e.g., Go read the book.)
- Polarity strength
- Target resolution
- Topic-specific predictiveness of features
- .....

Features

- Unigrams (presence vs. frequency)
- Higher-order n-grams (mixed results)
- Corpus frequency (Hapax legomena -- objective text is repeated)
- Part-of-speech (“love”--> love_NOUN)
- Position information (“good” --> good_END)
- Valence shifters (“don’t like” --> NOT_like)
- Target oriented features (“long” --> BATTERY_LIFE_long)
- Genre-specific features (“scary” --> HORROR_scary)
- Dependency parse features (http://nlp.stanford.edu:8080/parser/index.jsp)

Pang and Lee, EMNLP 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**

	Proposed word lists	Accuracy	Ties
Human 1	positive: *dazzling, brilliant, phenomenal, excellent, fantastic* negative: *suck, terrible, awful, unwatchable, hideous*	58%	75%
Human 2	positive: *gripping, mesmerizing, riveting, spectacular, cool, awesome, thrilling, badass, excellent, moving, exciting* negative: *bad, cliched, sucks, boring, stupid, slow*	64%	39%

Approaches

- Classification
- Regression
- Building genre-specific classifiers
- Inferring term-polarity with seeds/conjunctions (and, but)
  - Elegant, but over-priced; clever and informative
- Inferring labels heuristically (stars, emoticons)
- Self-training
- Co-training
- Domain adaptation

Domain Adaptation

- Challenges
  - Some features may not appear in the target domain
  - Some features may have the opposite polarity

Domain Adaptation

source domains

books

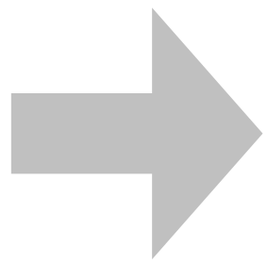
mobile phones

music albums

movies

laptops

restaurants



target domain

kitchen appliances

Related Tasks

- Detecting positive/negative judgement
- Predicting degree of positivity/negativity (regression)
- Extracting sentences that provide justification
- Extracting sentences that express comparison
- Predicting agreement/disagreement
- Viewpoint detection (pro vs. against)
- Detecting issue frames around debate