Finding Domain Specific Polar Words for Sentiment Classification

Mehrbod Sharifi

Outline

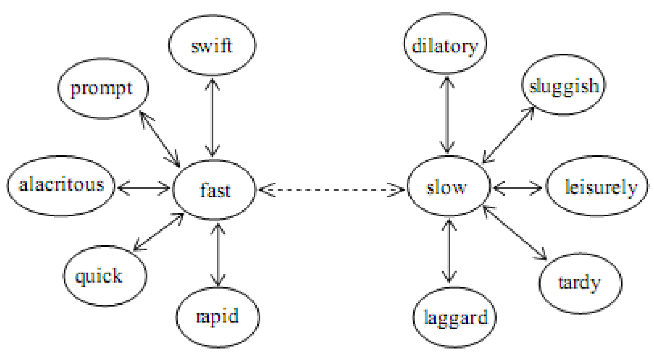
- Introduction
- Data & Previous Work
- Approach & Results
  - Prior Polarity Lexicons
  - Feature Extraction (Boosting)
  - Sequence Modeling (CRF)

Sentiment Classification

- Negative: "a lousy movie that's not merely unwatchable , but also unlistenable."
- Positive: "one of the greatest romantic comedies of the past decade."
- Negative: "these guys seem great to knock back a beer with but they're simply not funny performers."
- Negative: totally overwrought , deeply biased , and wholly designed to make you feel guilty about ignoring what the filmmakers clearly believe are the greatest musicians of all time."

Word Polarity

- Hatzivassiloglou & McKewon '97: Consider adjectives and extend by conjunctions (82%)
  - … simple and well-received …
  - … simplistic but well-received …
- Turney '02: $PMI(x,y) = \log_2 \frac{p(x,y)}{p(x)p(y)}$
  - SO(phrase) = PMI(phrase,"excellent")-PMI(phrase,"poor")
  - Estimate by web hits (74%)
- Liu '04: Start with seed sets and expand with WordNet



From Word to Sentences

- Minqing Hu and Bing Liu ’04: majority
- Kim & Hovy ‘04: product of sign, arithmetic or geometric mean (first and second was most useful)
- Popescu and Etzioni ‘05: Relaxation Labeling (optimization problem in three stages: word, phrase, sentence

Prior Polarity Lexicon
General Inquirer '00

- Manual

Entry	Source	Positiv	Negativ	Pstv	Affil	Ngtv	Hostile	Strong	Power
A	H4Lvd								
ABANDON	H4Lvd		Negativ			Ngtv			
ABANDONMENT	H4		Negativ						
ABATE	H4Lvd		Negativ						
ABATEMENT	Lvd								
ABDICATE	H4		Negativ						
ABHOR	H4		Negativ				Hostile		
ABIDE	H4	Positiv			Affil				
ABILITY	H4Lvd	Positiv						Strong	
ABJECT	H4		Negativ						
ABLE	H4Lvd	Positiv		Pstv				Strong	
ABNORMAL	H4Lvd		Negativ			Ngtv			
ABOARD	H4Lvd								
ABOLISH	H4Lvd		Negativ			Ngtv	Hostile	Strong	Power

http://www.wjh.harvard.edu/~inquirer/

Prior Polarity Lexicon
Subjectivity Clues (Riloff & Wiebe '03,'05)

- Automatically selected syntactic pattern

type=weaksubj len=1 word1=**abandoned** pos1=adj stemmed1=n priorpolarity=**negative**
type=weaksubj len=1 word1=**abandonment** pos1=noun stemmed1=n priorpolarity=**negative**
type=weaksubj len=1 word1=**abandon** pos1=verb stemmed1=y priorpolarity=**negative**
type=strongsubj len=1 word1=abase pos1=verb stemmed1=y priorpolarity=negative
type=strongsubj len=1 word1=abasement pos1=anypos stemmed1=y priorpolarity=negative
type=strongsubj len=1 word1=abash pos1=verb stemmed1=y priorpolarity=negative
type=weaksubj len=1 word1=abate pos1=verb stemmed1=y priorpolarity=negative
type=weaksubj len=1 word1=abdicate pos1=verb stemmed1=y priorpolarity=negative
type=strongsubj len=1 word1=aberration pos1=adj stemmed1=n priorpolarity=negative
type=strongsubj len=1 word1=aberration pos1=noun stemmed1=n priorpolarity=negative
type=strongsubj len=1 word1=abhor pos1=anypos stemmed1=y priorpolarity=negative
type=strongsubj len=1 word1=abhor pos1=verb stemmed1=y priorpolarity=negative
type=strongsubj len=1 word1=abhorred pos1=adj stemmed1=n priorpolarity=negative
type=strongsubj len=1 word1=abhorrence pos1=noun stemmed1=n priorpolarity=negative
type=strongsubj len=1 word1=abhorrent pos1=adj stemmed1=n priorpolarity=negative
type=strongsubj len=1 word1=abhorrently pos1=anypos stemmed1=n priorpolarity=negative
type=strongsubj len=1 word1=abhors pos1=adj stemmed1=n priorpolarity=negative
type=strongsubj len=1 word1=abhors pos1=noun stemmed1=n priorpolarity=negative

http://www.cs.pitt.edu/mpqa/

Prior Polarity Lexicon
SentiWordNet (Esuli & Sebastiani '06)

- Classification of word gloss in dictionary

# POS	offset	PosScore	NegScore	SynsetTerms
a	1000003	0.0	0.125	form-only#a#1
a	1000159	0.25	0.0	dress#a#1 full-dress#a#1
a	1000307	0.0	0.0	titular#a#5 nominal#a#6
a	1000440	0.0	0.0	prescribed#a#4 positive#a#5
a	1000554	0.0	0.25	perfunctory#a#2 pro_forma#a#1
a	1000681	0.0	0.5	semiformal#a#1 black-tie#a#1 semi-forma
a	10007	0.0	0.625	abstentious#a#1 abstinent#a#1
a	1000859	0.0	0.0	starchy#a#2 buckram#a#1 stiff#a#4
a	1001035	0.125	0.375	white-tie#a#1
a	1001157	0.0	0.0	informal#a#1
a	100126	0.5	0.0	viable#a#2
a	1001456	0.375	0.125	casual#a#3 everyday#a#2
a	1001581	0.0	0.0	free-and-easy#a#1 casual#a#8
a	1001755	0.0	0.375	folksy#a#2
a	1001882	0.0	0.625	unceremonious#a#1 unceremonial#a#1
a	1002013	0.0	0.25	formal#a#3
a	1002315	0.0	0.0	literary#a#3
a	1002508	0.0	0.0	informal#a#3
a	100261	0.0	0.125	vital#a#4
a	1002760	0.0	0.0	conversational#a#1 colloquial#a#1
a	1003005	0.0	0.0	vulgar#a#3 vernacular#a#1 common#a#5
a	1003296	0.0	0.0	epistolary#a#1 epistolatory#a#1
a	1003509	0.375	0.125	slangy#a#1
a	1003665	0.125	0.5	subliterary#a#1
a	1003815	0.25	0.375	unliterary#a#1 nonliterary#a#1
a	100393	0.0	0.75	dead#a#1

http://sentiwordnet.isti.cnr.it/

Datasets

- Pang & Lee (‘02,’04)
  - Polarity Dataset– Long reviews: 1K+, 1K- (avg. 780 words)
  - Subjectivity Dataset – Short review: 5K+, 5K- (avg. 21 words)
- Restaurant Reviews (50K+, 1-5 rating, avg. 34 words)

To use later:

- Wiebe '06 (MPQA)
- Liu '04
- TREC Blog '06

Pang ‘02

	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%

Figure 1: Baseline results for human word lists. Data: 700 positive and 700 negative reviews.

	Proposed word lists	Accuracy	Ties
Human 3 + stats	positive: *love, wonderful, best, great, superb, still, beautiful* negative: *bad, worst, stupid, waste, boring, ?, !*	69%	16%

	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**

Experiment (Accuracy reported)

- SVM: Short reviews: 74% - Long reviews: 81%
  - Boolean feature vector (tf or tfidf is worse)
  - Only selected feature with tf ≥ 4 (50% accuracy using all features)
  - Encoding negation was not helpful
- Boosting: 76%
  - Same for 1000-5000 rounds of training
  - Almost the same for long and short reviews
  - 5-gram lowered 3%

SVM Error Analysis

- Sequence Issues (e.g., … but …)
  - interesting , but not compelling .
  - the effort is sincere and the results are honest , but the film is so bleak that it's hardly watchable .
- Neg: not once does it come close to being exciting .
- Pos: while not all that bad of a movie , it's nowhere near as good as the original .

Selecting Polar Words - Example

Positive score, Negative score[,mutiplicity]:

- **GI** `the story is far-flung , illogical[0,1,1] , and plain[4,0,4] stupid[0,3,3] .  =Total=> 4 4`
- **SW** `the story is far-flung , illogical[0.625,0.375,2] , and plain[2.625,3.125,13] stupid[0.25,0.5,4] .  =Total=> 3.5 4`
- **SC** `the story is far-flung , illogical[0,1,1] , and plain[1,0,1] stupid[0,1,1] .  =Total=> 1 2`

Accuracy

- Baseline: Random or all in one class: 50%
- Short reviews
  - SC 70.9% (18.5% tie)
  - GI 70.4% (23.2% tie)
  - SW 59.9% (2.4% tie)
- Long reviews (tie <1%)
  - SC: 61.0%
  - GI: 56.6%
  - SW: 56.2%

6-10 times more error for negatives than positives

Feature Selection - Boosting

$$H(x) = sign\left(\sum_t \alpha_t h_t(x)\right) \qquad \alpha_t = \frac{1}{2}\ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right) > 0$$

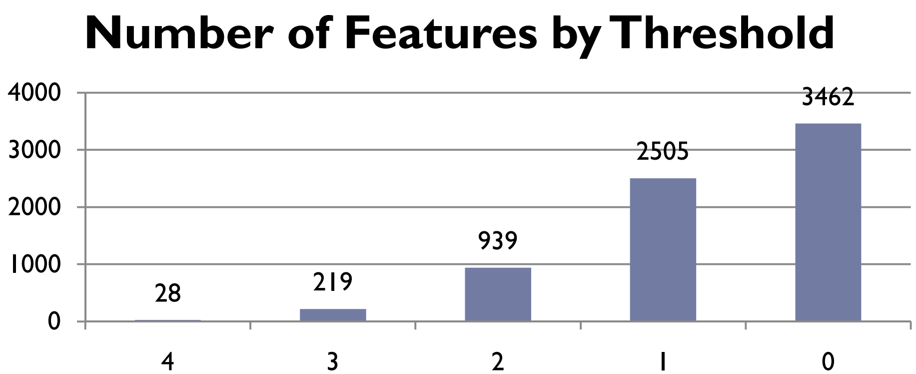
Unigram - Short reviews			
Neg		Pos	
bore	6.1	**outlandish**	5.7
dogs	4.7	moodiness	5.5
blank	4.7	liberating	4.9
stunt	4.7	combine	4.6
disappointment	4.6	**shrek**	4.5
benigni	4.6	screams	4.5
brawny	4.6	fulfill	4.4
whiny	4.5	**eyerolling**	4.4
gotten	4.5	tape	4.3
stumble	4.5	priceless	4.2
dimwitted	4.4	groantoguffaw	4.2
demmes	4.4	mesmerizing	3.8
limitations	4.3	concern	3.8
claim	4.3	vividly	3.7
mud	4.3	**bourne**	3.7
routine	4.3	glorious	3.7
paint	4.2	sly	3.7
disguise	4.0	ingenious	3.7
erratic	3.9	refreshingly	3.7
pointless	3.8	bride	3.7

Unigram - Short reviews (after pruning)			
Neg		Pos	
blank	4.7	moodiness	5.5
stunt	4.7	combine	4.6
disappointment	4.6	fulfill	4.4
brawny	4.6	priceless	4.2
whiny	4.5	mesmerizing	3.8
stumble	4.4	concern	3.8
claim	4.3	vividly	3.7
mud	4.3	glorious	3.7
routine	4.3	sly	3.7
disguise	3.9	ingenious	3.7
erratic	3.9	refreshingly	3.7
pointless	3.8	engrossing	3.6
incoherent	3.8	happily	3.6
horrible	3.7	accurate	3.6
uninspired	3.7	harrowing	3.5
choppy	3.6	gently	3.5
bother	3.6	image	3.5
exhausting	3.6	wash	3.4
strained	3.6	soulful	3.4
soggy	3.6	higher	3.4

Prune = overlap with prior priority

Feature Selection - Boosting

the story is far-flung , illogical , and plain[0,1.73,1] stupid[0,2.56,1] . =Total=> 0 4.30



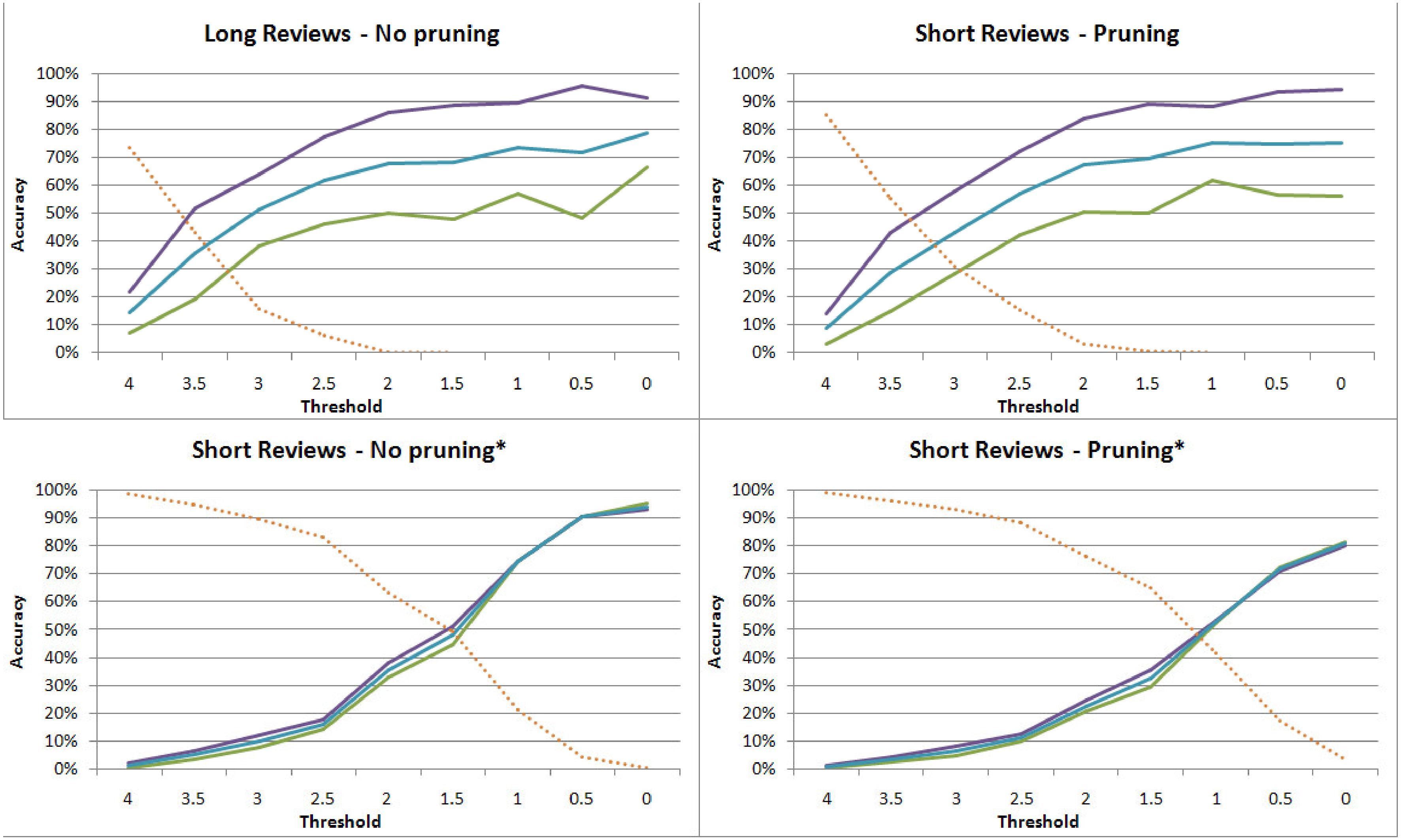
>3: will grab your children by the imagination and amaze them and amuse them . **0 0**

>2: will grab your children by the imagination and amaze[2.14,0.00] them and amuse them . **2.14 0**

>1: will grab your children by the imagination and amaze[2.14,0.00] them and amuse[1.01,0.00] them . **3.15 0**

>0: will grab your children by[0.00,0.11] the imagination and[0.23,0.00] amaze[2.14,0.00] them and[0.23,0.00] amuse[1.01,0.00] them . **3.60 0.11**

Feature Selection - Boosting



* Boosting is trained on 2/3 of this data

Sequence Model - CRF

- Setup
  - Boosting features threshold at 1 and pruned
  - 6 gram for CRF
  - Only subjective sentences
- Trained on 2/3 and tested on 1/3
  - repeated but not cross validation yet
  - Total tokens 59446 in test set
- Result
  - 56 + and 97 - more than training
  - 38 + and 29 – completely new
- Incorrect:
  - True - : missed 251 times and as + : 5
  - True + : missed 205 times and as - : 1
  - Correct: 3787 - and 3275 +
- However, classification improvement is small (added with weight 3)

+	-
fascinated	sorrowful
superlative	uninhibited
amazingly	irreverent
workmanlike	unfree
inventively	deserved
resent	minkoff
superstitious	dispossessed
funny/gritty	discomfort
densely	woodland
ethereal	missive

References

- Bo Pang and Lilian Lee, L., and Shivakumar Vaithyanathan. 2002 Thumbs up? Sentiment Classification using Machine Learning Techniques, *Proceedings of Empirical Methods in Natural Language Processing*
- Peter D. Turney and Michael L. Littman. 2002. Unsupervised learning of semantic orientation from a hundred-billion-word corpus. Technical Report EGB-1094, National Research Council Canada.
- Ellen Riloff, Janyce Wiebe, and Theresa Wilson. 2003 Learning Subjective Nouns Using Extraction Pattern Bootstrapping. Seventh Conference on Natural Language Learning (CoNLL-03). ACL SIGNLL. Pages 25-32.
- Ellen Riloff and Janyce Wiebe. 2003 Learning Extraction Patterns for Subjective Expressions. Conference on Empirical Methods in Natural Language Processing (EMNLP-03). ACL SIGDAT. Pages 105-112. Here are presentation slides.
- Theresa Wilson , Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. HLT-EMNLP-2005. Here are presentation slides**.**
- V.Hatzivassiloglou, K.McKeown. 1997. Predicting semantic orientation of adjectives.In ACL/EACL,pages174–181.
- M.Hu and B.Liu 2004. Mining and Summarizing Customer Reviews. In KDD, pages168–177, Seattle, WA.
- Soo-Min Kim and Eduard Hovy. 2004. Determining the Sentiment of Opinions. Proceedings of the COLING conference, Geneva, 2004
- Andrea Esuli and Fabrizio Sebastiani. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In Proceedings of LREC-06, 5th Conference on Language Resources and Evaluation, Genova, IT, 2006
- Ana-Maria Popescu and Oren Etzioni. 2005. Extracting Product Features and Opinions from Reviews, Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing 2005