Linguistic Expressions of Sentiment, Subjectivity & Stance

Ling575
Sentiment
April 1, 2014

Roadmap

- Motivation:
  - Why sentiment? Why now?
- A Word on Terminology
- Applications
- Challenges
- Approaches: Starting with the basics
  - Word level approaches to Polarity
- Course Mechanics

Why Sentiment?

- Plays a key role in decision-making
- We've always wondered "What do other people think?"
  - Ask friends for recommendations
  - Ask employers/landlords for references
  - Check with *Consumer Reports,* BBB, newspapers, etc
- What makes the Web different?
  - Access to enormous numbers of reviews, opinions
    - Largely unknown, non-expert
    - Widely accessible
  - Increasing numbers write reviews, blogs, opinions, etc

Why Sentiment?

- Plays a key role in decision-making
- We've always wondered "What do other people think?"
  - Ask friends for recommendations
  - Ask employers/landlords for references
  - Check with *Consumer Reports,* BBB, newspapers, etc
- What makes the Web different?
  - Access to enormous numbers of reviews, opinions
    - Largely unknown, non-expert
    - Widely accessible
  - Increasing numbers write reviews, blogs, opinions, etc

Why Sentiment?

- Plays a key role in decision-making
- We've always wondered "What do other people think?"
  - Ask friends for recommendations
  - Ask employers/landlords for references
  - Check with *Consumer Reports*, BBB, newspapers, etc
- What makes the Web different?
  - Access to enormous numbers of reviews, opinions
    - Largely unknown, non-expert
    - Widely accessible
  - Increasing numbers write reviews, blogs, opinions, etc

Why Sentiment?

- Plays a key role in decision-making
- We've always wondered "What do other people think?"
  - Ask friends for recommendations
  - Ask employers/landlords for references
  - Check with *Consumer Reports,* BBB, newspapers, etc
- What makes the Web different?
  - Access to enormous numbers of reviews, opinions
    - Largely unknown, non-expert
    - Widely accessible
  - Increasing numbers write reviews, blogs, opinions, etc

Frequency, Ubiquity & Impact

- Surveys say … (from Pang & Lee, 2008)
- Users
  - 81% have done online research for produce/service
    - 20% daily
  - 73-87% of readers report influenced by reviews
    - Will pay 20-99% more for 5* product than 4*
  - 30% research political issues: pro, con, endorsements
  - However, ~60% say: confusing, missing, overwhelming

Organizational Perspectives

- Vendors
  - Can gain access to quantities of info about products
  - However, sources diverse, fragmented, overwhelming

- eGov: Governmental eRulemaking Initiatives:
  - (www.regulations.gov)
  - Solicit direct citizen input on rules & regs
  - 400,000 comments received on single organic food labeling rule
  - Automatic tools crucial for coping with flood

Opinion Search

- Steps for a basic application
  - 1) Standard document retrieval search
    - Possibly with keywords like ‘reviews’, ’opinions’
  - 2) Identify review/opinionated portions of documents
    - Easy: Amazon, Yelp, etc
    - Harder: Blogs: often subjective, but highly varied, sloppy
  - 3) Identify expressed sentiment
    - Overall: Positive/negative review; 5*
    - Specific: opinions re features/aspects
  - 4) Summarization review content: scores, pros/cons,etc
- We’ll cover 2, 3, 4

Sentiment Explosion

- Early work on beliefs and metaphor
- 1994: early work on subjectivity (Wiebe)
  - Contrast: objective vs subjective content
- 2001: Huge increase in sentiment-related word
  - Why?
    - Development of machine learning techniques
    - Data availability: review aggregation sites
    - Awareness of intellectual, commercial opportunities

A Word on Terminology

- Explosion of research, explosion of terms
- *Subjectivity*: (Wiebe, 1994, and followers)
  - Motivated by Quirk's idea of "private state"
    - Opinions, evaluations, emotion, etc
  - Main goal: Distinguish subjective from objective
- *Affective Computing:*
  - Recognizing, synthesizing emotion content: happy, angry, sad, ...
- *Opinion mining*: Dave et al, '03
  - Search community: aggregate views of aspects of items
- *Sentiment analysis*: Chen & Das '01; Pang & Lee, '02
  - NLP community: initially polarity classification, now any

Applications

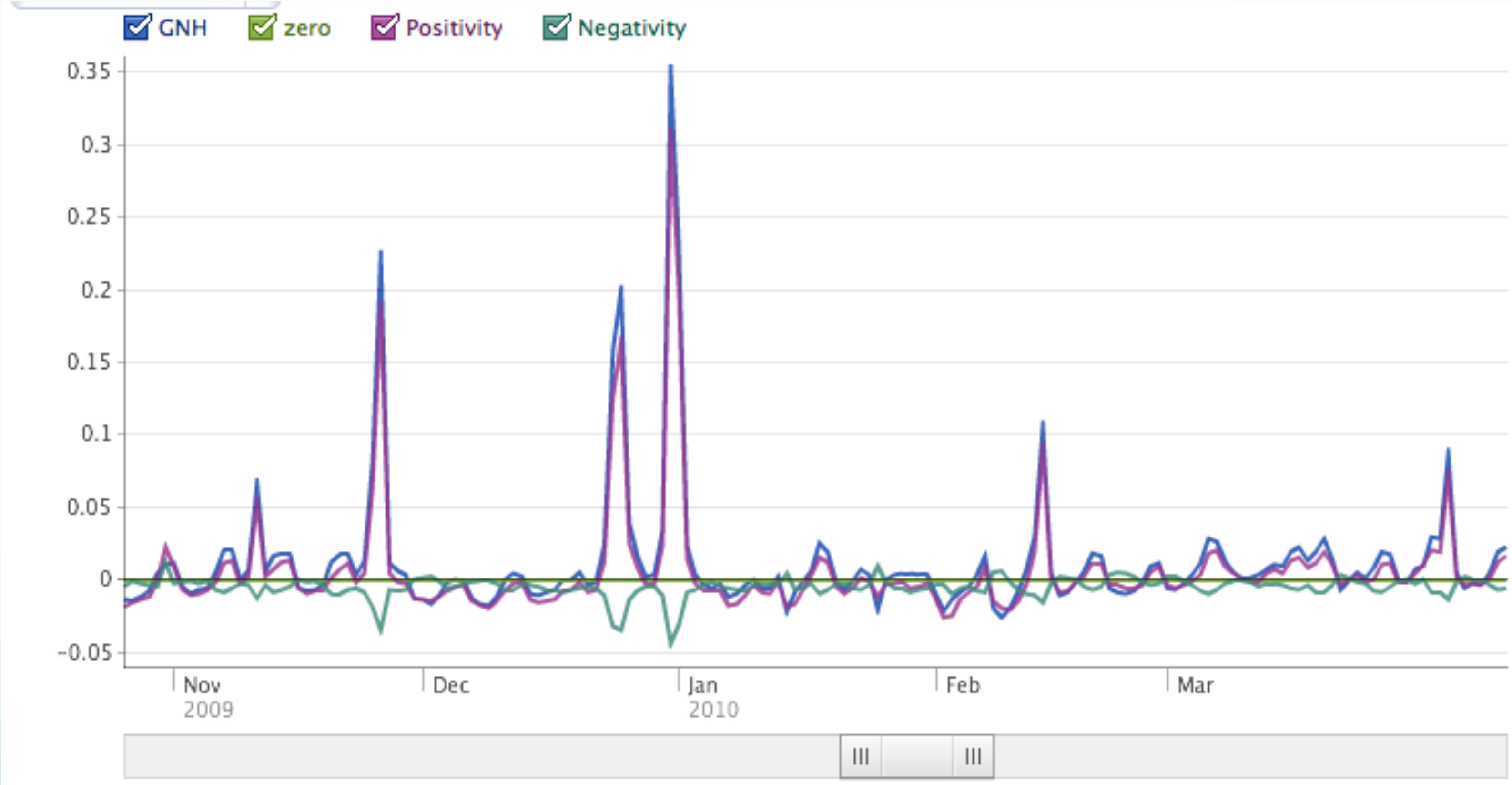
- Review sites:
  - Automation, aggregation, summarization
  - Verification: i.e. matching * ratings to text
- Component technology for:
  - Flame detection, question-answering, citation analysis
- Business intelligence:
  - Extract, summarize opinions about products, etc
- Tracking:
  - Political stances, depression in tweets, eGov

Applications: Google Product Search

From R. Feldman, 2013

Applications

- From C. Potts



- Figure: Facebook's Gross National Happiness interface (defunct?). Holidays register large happiness spikes. The happiness dips in January correspond roughly with the earthquake in Haiti (Jan 12) and its most serious aftershock (Jan 20).

Applications

- From C. Potts

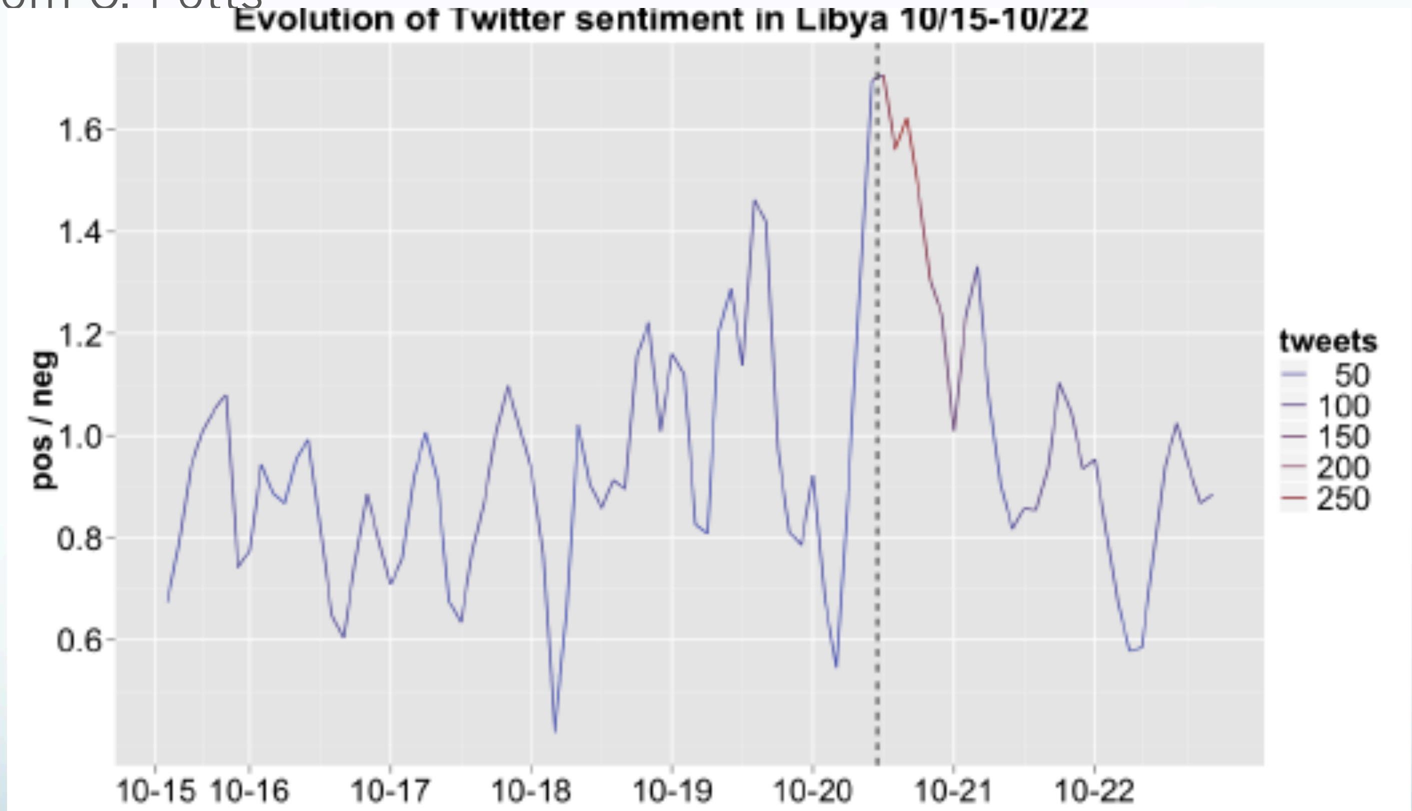


Figure: Twitter sentiment in tweets about Libya, from the project 'Modeling Discourse and Social Dynamics in Authoritarian Regimes'. The vertical line marks the timing of the announcement that Gaddafi had been killed.

Other Applications

- “Twitter mood predicts the stock market”
  - Bollen et al, 2010
- “Predicting Postpartum Changes in Emotion and Behavior via Social Media”
  - M. De Choudhury et al, 2013
- "Flaming drives online social networks”
  - Condliffe, 2010
- “Get out the vote: Determining support or opposition from Congressional floor-debate transcripts”
  - Thomas et al.

Situating Sentiment

- Text classification:
  - Typically assigns documents to finite set of categories
    - Potentially large #, generally unrelated/disjoint
  - Sentiment: very small # of categories, opposing/scale
- Information extraction:
  - Automatically fill information slots in template via text
    - Templates highly variable, specific to domain
  - Sentiment analysis fills fixed fields across domains
    - Holder, type, strength, target

Solving Sentiment

- Basic task: Polarity classification
  - Label subjective unit as positive or negative
- Example: “The most thoroughly joyless and inept film of the year, and one of the worst of the decade” [Mick LaSalle, of *Gigli*][via L. Lee, 2008]
  - Thumbs up or down??
- Easy, right? Why?
- Obvious lexical polarity indicators:
  - Worst !! , also joyless, inept

Is it that easy?

- Just pick words associated with positive/negative
- Human word picking experiment

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

Is it that easy?

- Just pick words associated with positive/negative
- Human word picking experiment

	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%

Is it that easy?

- Just pick words associated with positive/negative
- Human word picking experiment

	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%
Statistics-based	positive: *love, wonderful, best, great, superb, still, beautiful* negative: *bad, worst, stupid, waste, boring, ?, !*		

Is it that easy?

- Just pick words associated with positive/negative
- Human word picking experiment

	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%
Statistics-based	positive: *love, wonderful, best, great, superb, still, beautiful* negative: *bad, worst, stupid, waste, boring, ?, !*	69%	16%

- Picking the right words is hard: non-obvious, domain dependent

When cue words fail...

- Let's just use 'great'
  - This laptop is a great deal.
  - A great deal of media attention surrounded the release of the new laptop.
  - This laptop is a great deal...and I've got a nice bridge you might be interested in.

Example from L. Lee, 2008

Finding the right words

- Sometimes there are no overt sentiment words
  - Subtle, indirect
    - "She runs the gamut of emotions from A to B."
  - (Due to Bob Bland.)
    - "Go read the book." In a book review
      - Vs
    - "Go read the book." In a movie review
    - Context dependent
  - This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However it can't hold up.
    - Order dependent

Confounds

- Many factors influence interpretation of sentiment:
  - Lexical content
    - Specific: sentiment dictionaries
    - General: classifiers over unigrams can reach 80%
  - Order
  - Context: Linguistics or real-world
  - Negation: That is not a book I want to read.
  - Syntax: A is better than B vs B is better than A.
  - Discourse relations
  - Domain: ‘unpredictable’: good in a story, bad in steering

Confounds

- Many factors influence interpretation of sentiment:
  - Lexical content
    - Specific: sentiment dictionaries
    - General: classifiers over unigrams can reach 80%
  - Order
  - Context: Linguistics or real-world
  - Negation: That is not a book I want to read.
  - Syntax: A is better than B vs B is better than A.
  - Discourse relations
  - Domain: 'unpredictable': good in a story, bad in steering

Broader Questions

- How do expression and interpretation of sentiment differ
  - Across languages
  - Between monolog and dialog
  - Across registers: Editorials vs review sites vs Twitter
  - Between text and speech

Sentiment-Related Tasks

- *Subjectivity recognition*: objective/subjective
  - There is a tree on the corner. I think it's beautiful.
- *Polarity classification*: Ling. unit positive/negative
  - The film was great. No, it was dreadful.
- *Aspect analysis*: +/- of specific features, contribution
  - The phone is great, but the screen is very flimsy.
- *Sentiment spam recognition*: Finding fake reviews
- *Domain adaptation*: applying models across domains

Polarity Classification

- Straightforward task:
  - Given a linguistic unit, determine positive/negative
    - Maybe rating: 1-5 stars
  - Most tasks assume subjective text, exclude neutral

- Can provide insight into:
  - Resources
  - Features
  - Models

Baseline Approaches

- Early approaches: Intuitive
  - Use lexicon of positive/negative words
  - Heuristic:
    - Count: |P| = # positive terms, |N| = # negative terms
    - If |P| > |N|, assign positive, else negative
  - Simple!
  - Can work surprisingly well!
- Problems?

Issues with Lexicon Approaches

- Where does the list come from?
- Does the same list work across
  - Domains, registers, language types,…?
- How good/complete is the lexicon?
- Too simple: easily misled
  - Negation, order, overall vs local, etc

Sentiment Lexicons

- General Inquirer (Classic):
  - Words labeled: positiv/negativ; o.w.
    - Includes broad POS tags

	Entry	Positiv	Negativ	Hostile	. . . (184 classes)	Othtags	Defined
1	A					DET ART	. . .
2	ABANDON		Negativ			SUPV	
3	ABANDONMENT		Negativ			Noun	
4	ABATE		Negativ			SUPV	
5	ABATEMENT					Noun	
⋮							
35	ABSENT#1		Negativ			Modif	
36	ABSENT#2					SUPV	
⋮							
11788	ZONE					Noun	

Sentiment Lexicons

- MPQA: (Wiebe et al)
  - Designed for subjectivity analysis (~8200 wds)
    - Level of subjectivity, polarity, POS, etc.

1.	type=weaksubj	len=1	word1=abandoned	pos1=adj	stemmed1=n	priorpolarity=negative
2.	type=weaksubj	len=1	word1=abandonment	pos1=noun	stemmed1=n	priorpolarity=negative
3.	type=weaksubj	len=1	word1=abandon	pos1=verb	stemmed1=y	priorpolarity=negative
4.	type=strongsubj	len=1	word1=abase	pos1=verb	stemmed1=y	priorpolarity=negative
5.	type=strongsubj	len=1	word1=abasement	pos1=anypos	stemmed1=y	priorpolarity=negative
6.	type=strongsubj	len=1	word1=abash	pos1=verb	stemmed1=y	priorpolarity=negative
7.	type=weaksubj	len=1	word1=abate	pos1=verb	stemmed1=y	priorpolarity=negative
8.	type=weaksubj	len=1	word1=abdicate	pos1=verb	stemmed1=y	priorpolarity=negative

Sentiment Lexicons

- Bing Liu's Opinion Lexicon
  - ~2000 positive words
  - ~4700 negative words
  - Includes slang, common misspellings,etc
- LIWC: (Linguistic Inquiry and Word Counts)
  - Proprietary but we have a copy
  - Categorized regular expressions

Affect	abandon*, abuse*, abusi*, accept, accepta*, accepted, accepting, accepts, ache*
Posemo	accept, accepta*, accepted, accepting, accepts, active*, admir*, ador*, advantag*
Negemo	abandon*, abuse*, abusi*, ache*, aching, advers*, afraid, aggravat*, aggress*,
Anx	afraid, alarm*, anguish*, anxi*, apprehens*, asham*, aversi*, avoid*, awkward*
Anger	jealous*, jerk, jerked, jerks, kill*, liar*, lied, lies, lous*, ludicrous*, lying, mad

Sentiment Lexicons

- SentiWordNet:
  - Augments WordNet entries with Pos/Neg *scores*

POS	ID	PosScore	NegScore	SynsetTerms	Gloss
a	00001740	0.125	0	able#1	(usually followed by 'to') having the necessary means or [. . .]
a	00002098	0	0.75	unable#1	(usually followed by 'to') not having the necessary means or [. . .]

Sentiment Lexicon Analysis

- Issues:
  - Not particularly large
  - Language, register sensitive
  - Different goals, labeling
    - Consistency?
    - Disagreement analysis (by C. Potts)

	MPQA	Opinion Lexicon	Inquirer	SentiWordNet	LIWC
MPQA	—	33/5402 (0.6%)	49/2867 (2%)	1127/4214 (27%)	12/363 (3%)
Opinion Lexicon		—	32/2411 (1%)	1004/3994 (25%)	9/403 (2%)
Inquirer			—	520/2306 (23%)	1/204 (0.5%)
SentiWordNet				—	174/694 (25%)
LIWC					—

Sentiment Lexicon Analysis

- Many issues still unresolved
- Possible solution for domain sensitivity:
  - Learn a lexicon for the relevant data
  - Range of approaches:
    - Unsupervised techniques
    - Domain adaptation
    - Semi-supervised methods
- However, still fundamentally limited

Machine Learning Baselines

- Similar to much of contemporary NLP
- Sentiment analysis explosion happened when
  - Large datasets of opinionated content met
  - Large-scale machine learning techniques
- Polarity classification as machine learning problem
  - Features?
  - Models?

Baseline Feature Extraction

- Basic text features?
  - Bag-of-words, of course
    - N-grams

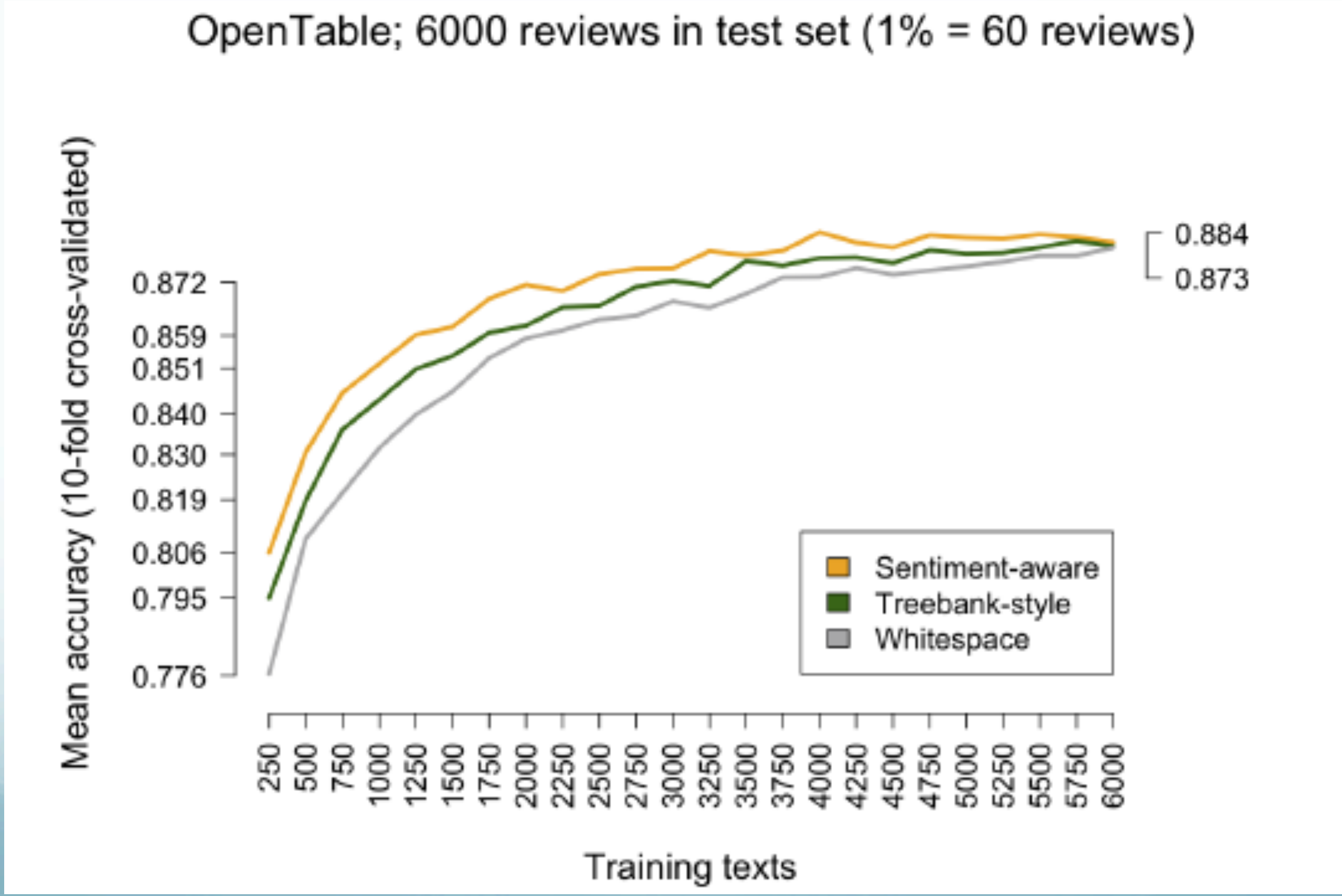
- Basic extraction:
  - Tokenization?
  - Stemming?
  - Negation?

Tokenizing

- Relatively simple for well-formed news
- Sentiment analysis needs to work on:
  - Sloppy blogs, tweets, informal material
  - What's necessary?
    - Platform markup handling/extraction
    - Emoticons ☺
    - Normalize lengthening
    - Maintain significant capitalization
    - Handle swear masks (e.g. %$^$ing)
- Comparisons on 12K OpenTable reviews: 6K: 4,5; 6K: 1,2
    - Results from C. Potts

Sentiment-Aware Tokenization

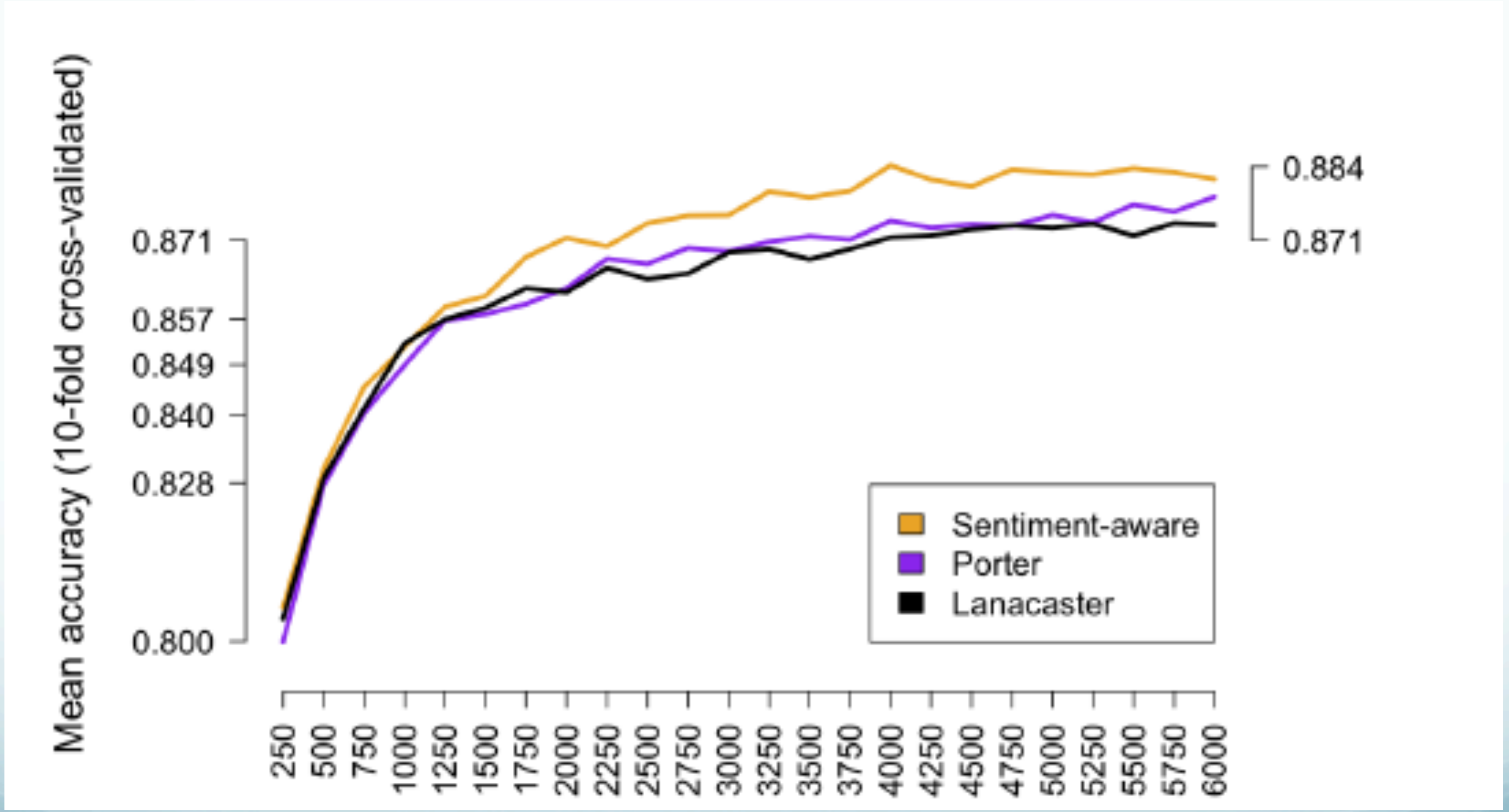
- From C. Potts



Stemming

- Should we stem?
  - Pros:
    - Reduces vocabulary, shrinks feature space
    - Removes irrelevant distinctions
  - Cons:
    - Can collapse relevant distinctions!

Stemming Impact on Sentiment Classification



Take home: Don't just grab a stemmer for sentiment analysis

Sentiment meets the Porter Stemmer

- Porter stemmer:
  - Classic heuristic rule cascade
    - Repeatedly strips off suffixes based on patterns
    - Highly aggressive
  - Applied to the General Inquirer
    - Destroys key contrasts

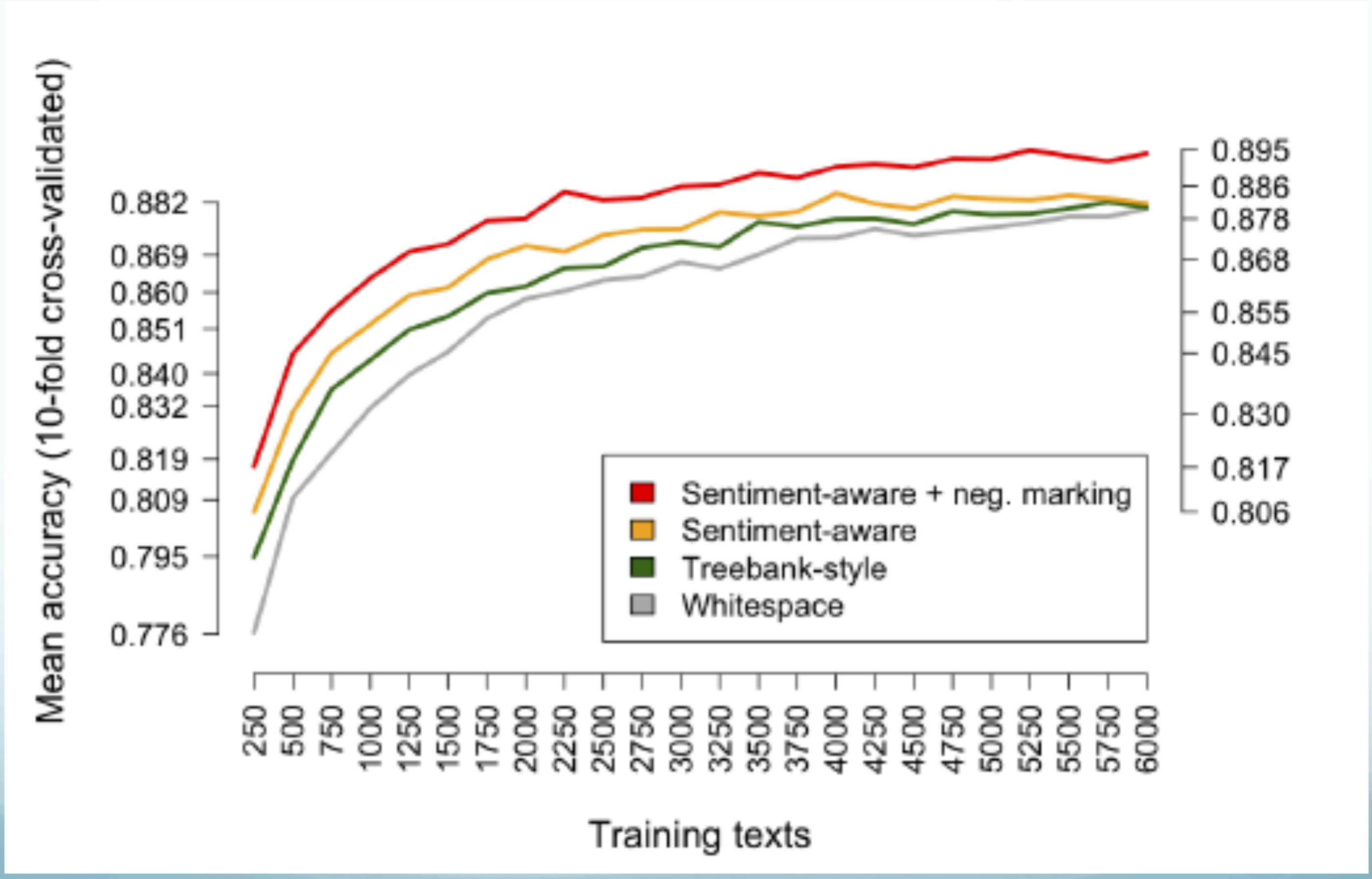
Positiv	Negativ	Porter stemmed
defense	defensive	defens
extravagance	extravagant	extravag
affection	affectation	affect
competence	compete	compet
impetus	impetuous	impetu
objective	objection	object

Naïve Negation Handling

- Negation:
  - The book was not good.
  - I did not enjoy the show.
  - No one enjoyed the movie.
- Approach due to Chen & Das, 2001
  - Add _NEG to each token between negation and end of clause punctuation
    - I did not enjoy the show. →
      - I did not enjoy_NEG the_NEG show_NEG

Impact of Negation Marking on Sentiment Analysis

- Even simple handling provides a boost



Bag-of-Words Representation

- Do polarity classification on:

Jane	so	want	from	over	that
can’t		beat	madden	shinbone	up
read	my	Austen	Prejudice	reader	her
frenzy	Pride	conceal		I	
and	books		Everytime	with	dig
the	own	skull		to	me

Bag-of-Words Representation

- Do polarity classification on:

Jane	so	want	from	over	that
can't		beat	madden	shinbone	up
read	my	Austen	Prejudice	reader	her
frenzy	Pride	conceal		I	
and	books		Everytime	with	dig
the	own	skull		to	me

Bag-of-Words Representation

- Do polarity classification on:

Jane	so	want	from	over	that
can’t		beat	madden	shinbone	up
read	my	Austen	Prejudice	reader	her
frenzy	Pride	conceal		I	
and	books		Everytime	with	dig
the	own	skull		to	me

Full text: Jane Austen’s book madden me so that I can’t conceal my frenzy from the reader. Everytime I read ‘Pride and Prejudice’ I want to dig her up and beat her over the skull with her own shinbone.

Bag-of-Words Representation

- Do polarity classification on:

Jane	so	want	from	over	that
can't		beat	madden	shinbone	up
read	my	Austen	Prejudice	reader	her
frenzy	Pride	conceal		I	
and	books		Everytime	with	dig
the	own	skull		to	me

Full text: Jane Austen's book madden me so that I can't conceal my frenzy from the reader. Everytime I read 'Pride and Prejudice' I want to dig her up and beat her over the skull with her own shinbone. - Mark Twain

Bag-of-Words

- Clearly, bag-of-words can not capture all nuances
  - Polarity classification hard for humans on that basis
- However, forms the baseline for many systems
- Can actually be hard to beat
  - MaxEnt classifiers with unigrams: >= 80%
    - On many polarity classification tasks
  - Current best results on polarity classification in dialog:
    - Combination of word, character, phoneme n-grams ~90% F-measure

Current Approaches

- Aim to improve over these baselines by
  - Better feature engineering
    - Modeling syntax, context, discourse, pragmatics
  - More sophisticated machine learning techniques
    - Beyond basic Naïve Bayes or MaxEnt models
- Recent state-of-the-art results (Socher et al)
  - Full parsing, syntactic analysis
  - Deep tensor network models

Course Mechanics

- Individual
  - Critical reading assignments
    - Weekly – one paper
- Groups of 2-3
  - Lead topic presentation/discussion: once
    - Select from list of topics, readings
    - Analyze, discuss in class
  - Term project
    - Explore specific topic in depth
      - Can implementation or analysis + write-up
      - Linguistics elective: talk to me

Datasets

- Diverse data sets:
  - Web sites: Lillian Lee's and Bing Liu's
- Movie review corpora
- Amazon product review corpus
- Online and Congressional floor debate corpora
- Multi-lingual corpora: esp. NTCIR
- MPQA subjectivity annotation news corpus