Sentiment and Opinion Mining

Learning About the World

Biases and Ethics

Yoav Goldberg

BIU NLP

Exam

- 3 hours (you may not need all of it)
- Open materials

Exam

- design a system to do X
- here is a system, how/why can it fail?
- here is an output of a system, which system is it?
- linguistic structures and annotations
- suggest features for a problem
- compute values according to METHODX
- fix this algorithm
- terms and concepts
- is X a good solution to Y? why? why not?

Information and Relation Extraction

- The most common / important need in NLP today.

Information and Relation Extraction

- There is never enough training data.
- Do you trust a computer to generalize well?
  - ...compared to the rules you write yourself?

Information and Relation Extraction

- Rule writing!

Information and Relation Extraction

- Rule based approach:
  - Transparent!
  - You know it will generalize well.
- But its really hard to write the rules... :((

Why is rule writing hard?

- Do we write them on top of text?
    - On top of parse trees?
        - On top of something else?
- What language do we use?

Why is rule writing hard?

- How do we come up with good rules?
  - When we see a sentence, we can write a rule.
    - But will it generalize?
    - And do we cover the corner cases?
  - What about the sentences we didn't see?

How can we improve this??

human in the loop / computer in the loop

How can we improve this??

human in the loop / computer in the loop

- How can we use Machine Learning to **help people** Write Good Rules?
- How can we use Machine Learning to **help people** produce good lexicons?

Sentiment Analysis and Opinion Mining

Sentiment Analysis

What is Sentiment Analysis?

Stanford/Coursera course, slides by Dan Jurafsky

Positive or negative movie review?

- unbelievably disappointing

- Full of zany characters and richly applied satire, and some great plot twists

- this is the greatest screwball comedy ever filmed

- It was pathetic. The worst part about it was the boxing scenes.

Dan Jurafsky

Google Product Search

HP Officejet 6500A Plus e-All-in-One Color Ink-jet - Fax / copier / printer / scanner
$89 online, **$100** nearby ★★★★☆ 377 reviews
September 2010 - Printer - HP - Inkjet - Office - Copier - Color - Scanner - Fax - 250 sh

Reviews

Summary - Based on 377 reviews

1 star | 2 | 3 | 4 stars | 5 stars

What people are saying

ease of use	"This was very easy to setup to four computers."
value	"Appreciate good quality at a fair price."
setup	"Overall pretty easy setup."
customer service	"I DO like honest tech support people."
size	"Pretty Paper weight."
mode	"Photos were fair on the high quality mode."
colors	"Full color prints came out with great quality."

Bing Shopping

HP Officejet 6500A E710N Multifunction Printer

Product summary   Find best price   **Customer reviews**   Specifications   Related items

$121.53 - $242.39 (14 stores)

Compare

Average rating (144)

5 stars	(55)
4 stars	(54)
3 stars	(10)
2 stars	(6)
1 star	(23)
0 stars	(0)

Most mentioned

Performance	(57)
Ease of Use	(43)
Print Speed	(39)
Connectivity	(31)

More ▼

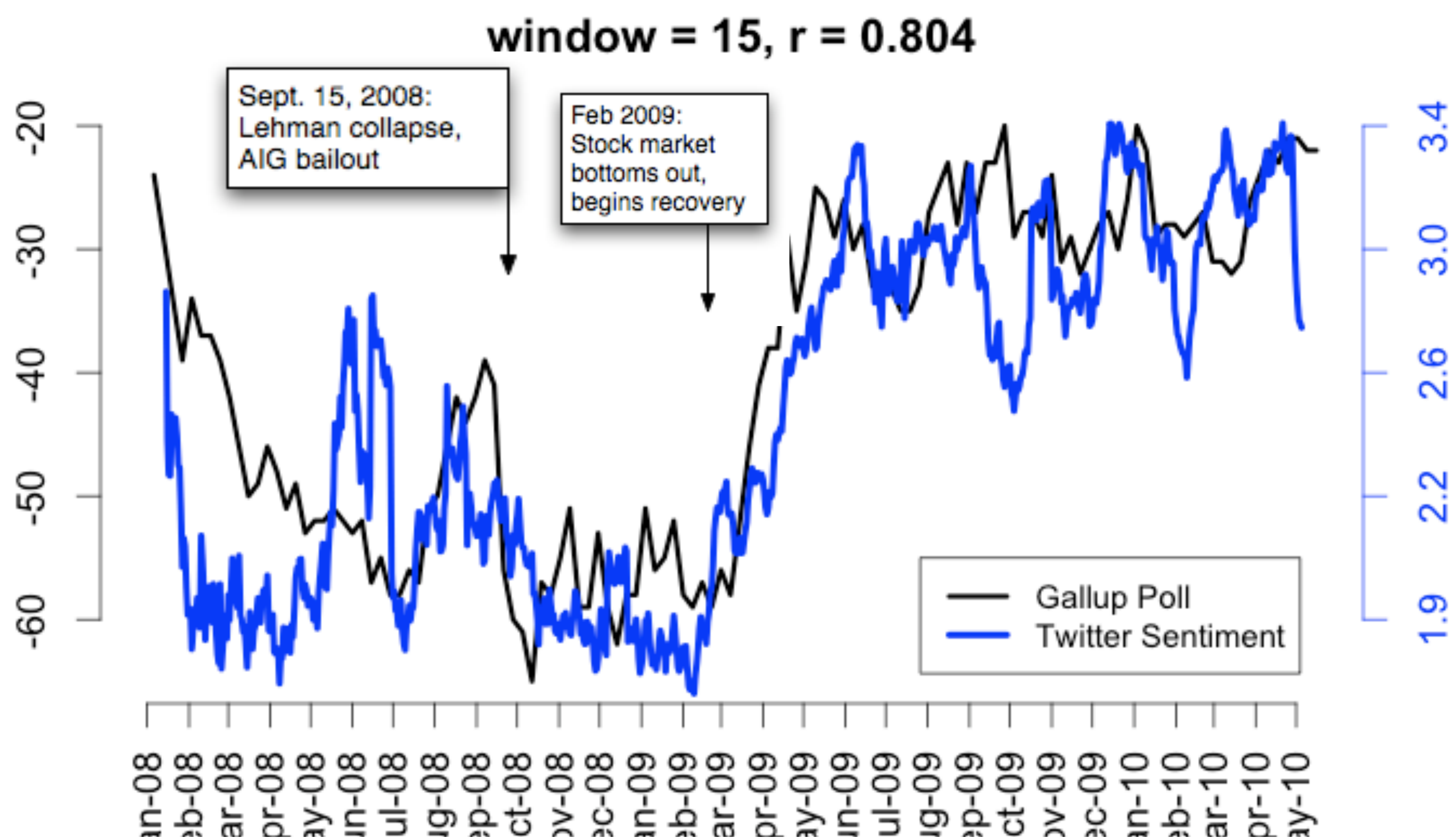
Show reviews by source

Best Buy (140)
CNET (5)
Amazon.com (3)

Dan Jurafsky

Twitter sentiment versus Gallup Poll of Consumer Confidence

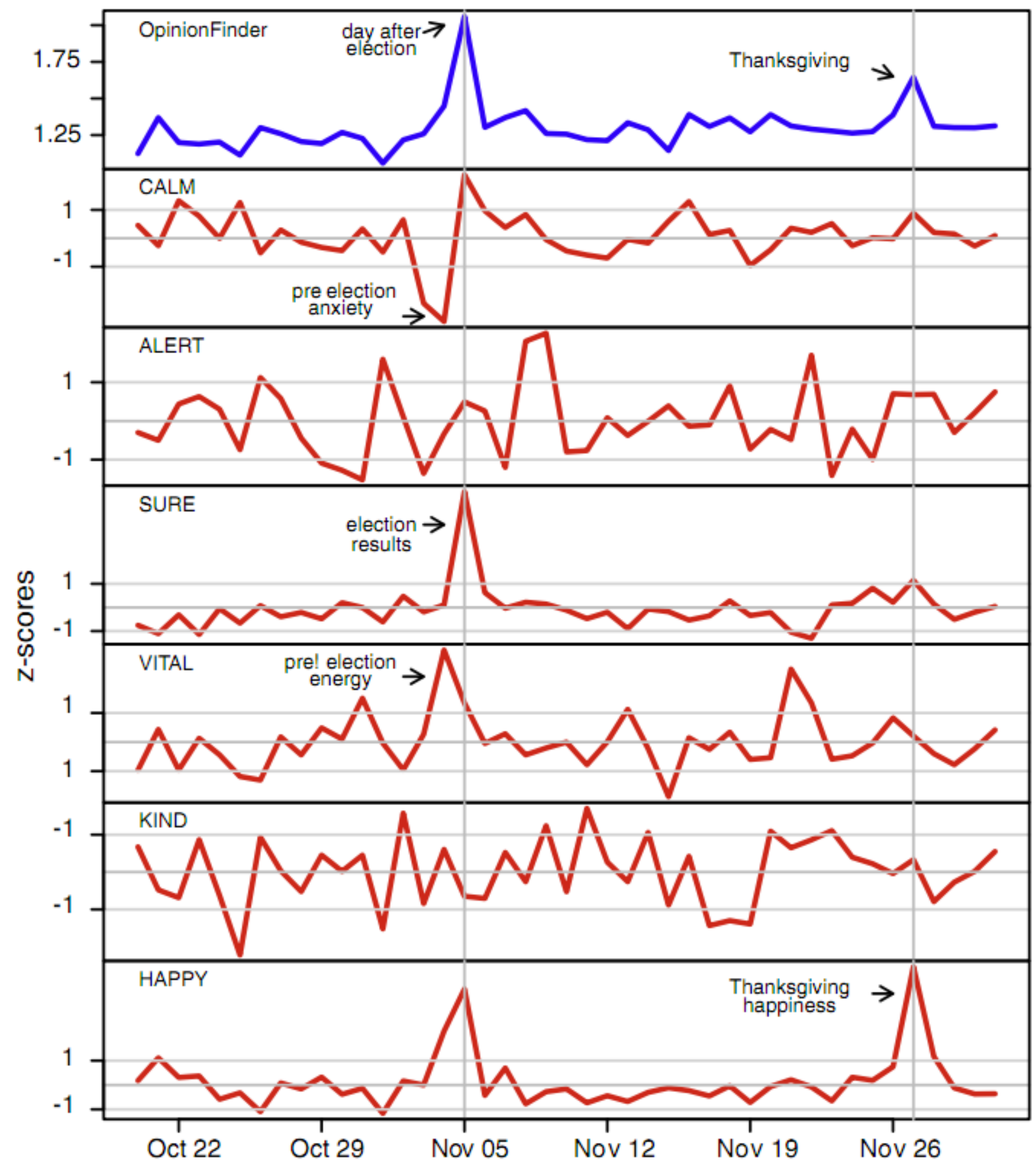
Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In ICWSM-2010



Twitter sentiment:

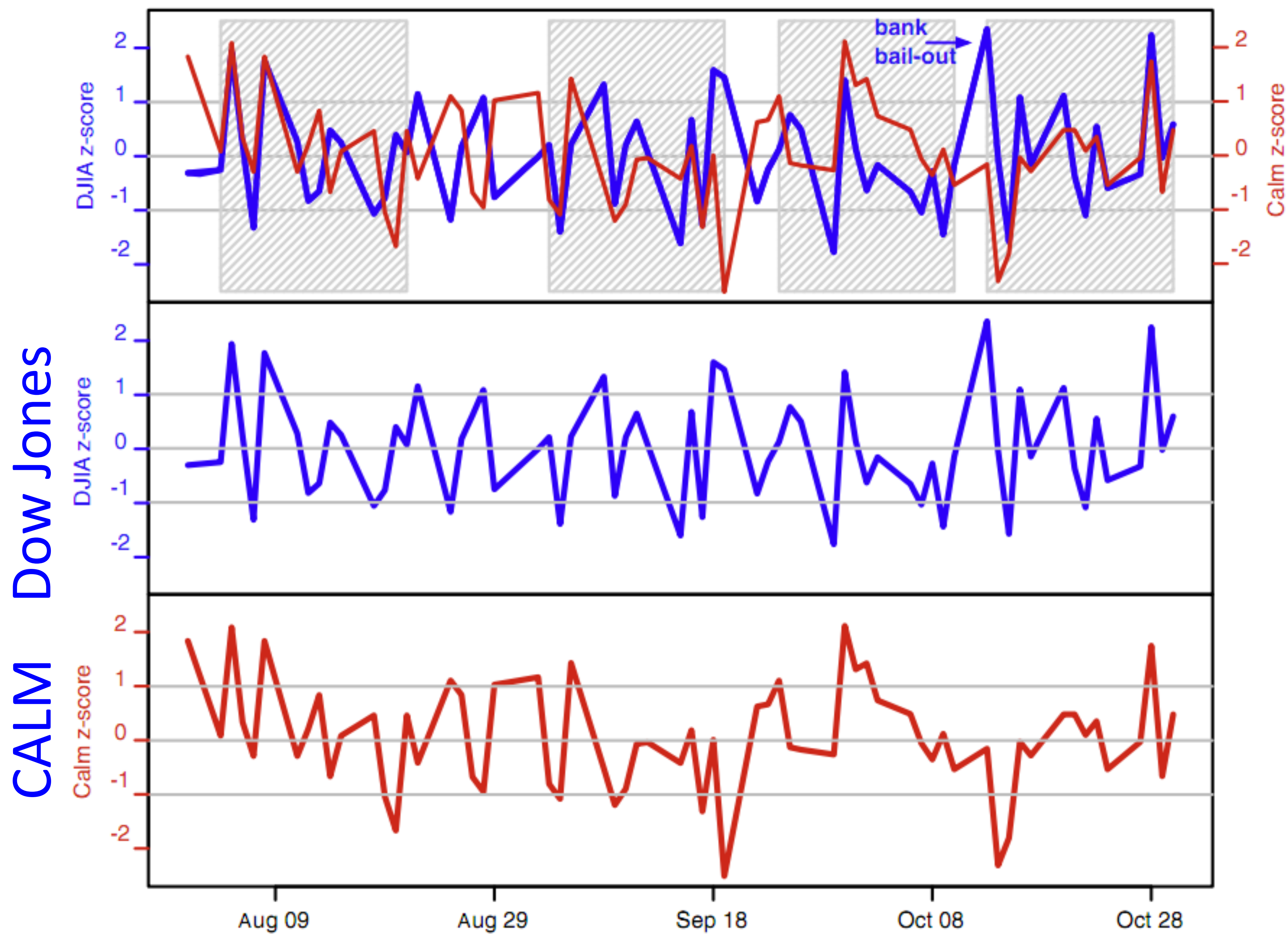
Johan Bollen, Huina Mao, Xiaojun Zeng. 2011. [Twitter mood predicts the stock market,]

Journal of Computational Science 2:1, 1-8. 10.1016/j.jocs.2010.12.007.



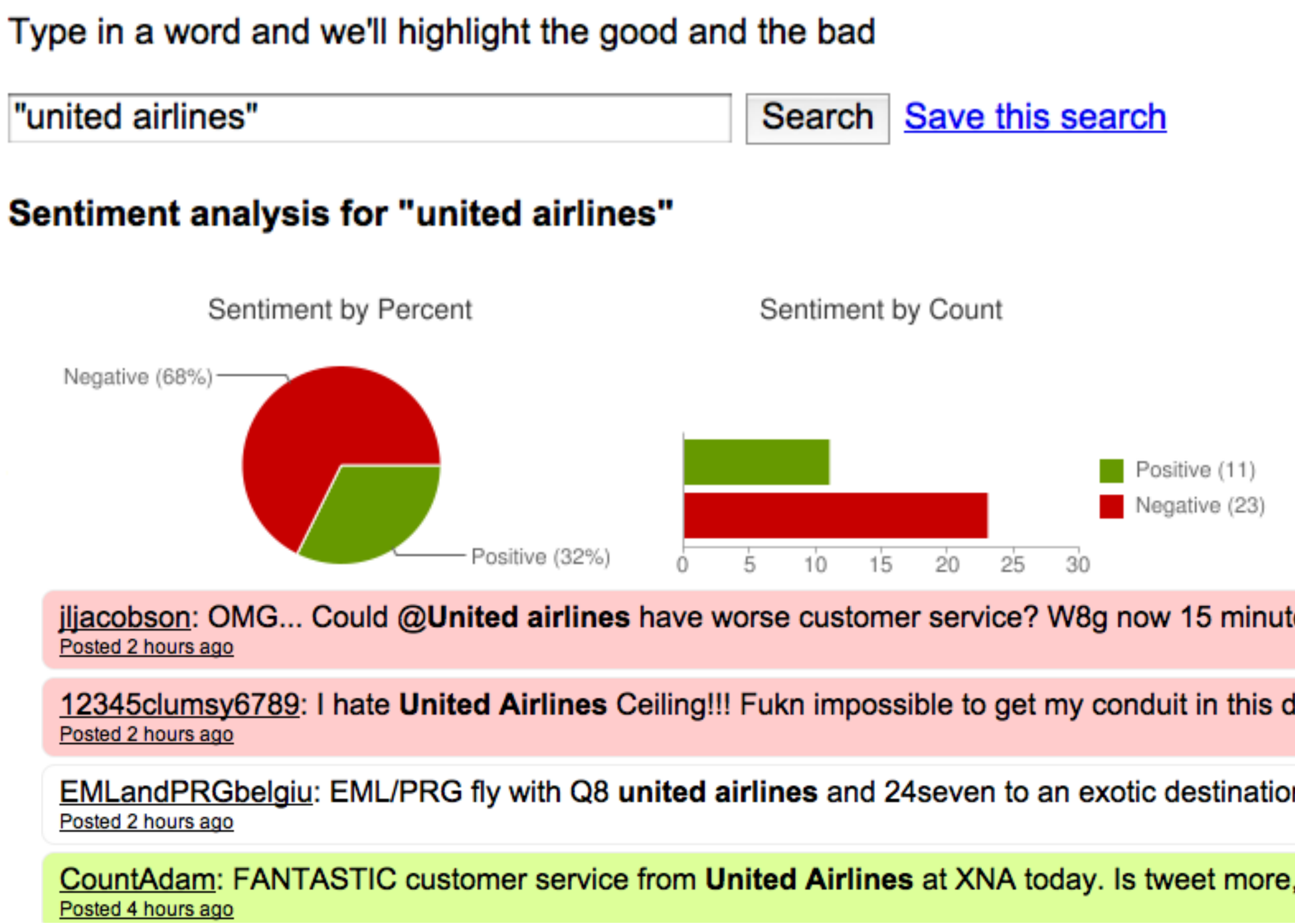
Bollen et al. (2011)

- CALM predicts DJIA 3 days later
- At least one current hedge fund uses this algorithm



Target Sentiment on Twitter

- Twitter Sentiment App
- Alec Go, Richa Bhayani, Lei Huang. 2009. Twitter Sentiment Classification using Distant Supervision



Sentiment analysis has many other names

- Opinion extraction
- Opinion mining
- Sentiment mining
- Subjectivity analysis

Why sentiment analysis?

- *Movie*: is this review positive or negative?
- *Products*: what do people think about the new iPhone?
- *Public sentiment*: how is consumer confidence? Is despair increasing?
- *Politics*: what do people think about this candidate or issue?
- *Prediction*: predict election outcomes or market trends from sentiment

Scherer Typology of Affective States

- **Emotion**: brief organically synchronized … evaluation of a major event
  - *angry, sad, joyful, fearful, ashamed, proud, elated*
- **Mood**: diffuse non-caused low-intensity long-duration change in subjective feeling
  - *cheerful, gloomy, irritable, listless, depressed, buoyant*
- **Interpersonal stances**: affective stance toward another person in a specific interaction
  - *friendly, flirtatious, distant, cold, warm, supportive, contemptuous*
- **Attitudes**: enduring, affectively colored beliefs, dispositions towards objects or persons
  - *liking, loving, hating, valuing, desiring*
- **Personality traits**: stable personality dispositions and typical behavior tendencies
  - *nervous, anxious, reckless, morose, hostile, jealous*

Scherer Typology of Affective States

- **Emotion**: brief organically synchronized … evaluation of a major event
  - *angry, sad, joyful, fearful, ashamed, proud, elated*
- **Mood**: diffuse non-caused low-intensity long-duration change in subjective feeling
  - *cheerful, gloomy, irritable, listless, depressed, buoyant*
- **Interpersonal stances**: affective stance toward another person in a specific interaction
  - *friendly, flirtatious, distant, cold, warm, supportive, contemptuous*
- **Attitudes: enduring, affectively colored beliefs, dispositions towards objects or persons**
  - *liking, loving, hating, valuing, desiring*
- **Personality traits**: stable personality dispositions and typical behavior tendencies
  - *nervous, anxious, reckless, morose, hostile, jealous*

Sentiment Analysis

- Sentiment analysis is the detection of **attitudes**

  "enduring, affectively colored beliefs, dispositions towards objects or persons"

  1. **Holder (source)** of attitude
  2. **Target (aspect)** of attitude
  3. **Type** of attitude
     - From a set of types
       - *Like, love, hate, value, desire,* etc.
     - Or (more commonly) simple weighted **polarity**:
       - *positive, negative, neutral,* together with *strength*
  4. **Text** containing the attitude
     - Sentence or entire document

Sentiment Analysis

- Simplest task:
  - Is the attitude of this text positive or negative?
- More complex:
  - Rank the attitude of this text from 1 to 5
- Advanced:
  - Detect the target, source, or complex attitude types

Sentiment Analysis Symposium 2014

Practical Sentiment Analysis Tutorial

Jason Baldridge

@jasonbaldridge

Associate Professor

THE UNIVERSITY OF TEXAS AT AUSTIN

Co-founder & Chief Scientist

people pattern

Polarity classification [slide from Lillian Lee]

- Consider just classifying an avowedly subjective text unit as either positive or negative (“thumbs up or “thumbs down”).
- One application: review summarization.
    - Elvis Mitchell, May 12, 2000: *It may be a bit early to make such judgments, but <u>Battlefield Earth</u> may well turn out to be the worst movie of this century*.
- Can’t we just look for words like “great”, “terrible”, “worst”?
- Yes, but ... learning a sufficient set of such words or phrases is an active challenge.

Using a lexicon [slide from Lillian Lee]

- From a small scale human study:

	Proposed word lists	Accuracy
Subject 1	Positive: dazzling, brilliant, phenomenal, excellent, fantastic Negative: suck, terrible, awful, unwatchable, hideous	58%
Subject 2	Positive: gripping, mesmerizing, riveting, spectacular, cool, awesome, thrilling, badass, excellent, moving, exciting Negative: bad, cliched, sucks, boring, stupid, slow	64%
Automatically determined (from data)	Positive: love, wonderful, best, great, superb, beautiful, still Negative: bad, worst, stupid, waste, boring, ?, !	69%

Polarity words are not enough [slide from Lillian Lee]

- Can't we just look for words like "great" or "terrible"?
- Yes, but ...
  - This laptop is <u>a great deal</u>.
  - <u>A great deal</u> of media attention surrounded the release of the new laptop.
  - This laptop is <u>a great deal</u> ... and I've got a nice bridge you might be interested in.

Polarity words are not enough

- **Polarity flippers**: some words change positive expressions into negative ones and vice versa.
  - **Negation**: *America still needs to be focused on job creation.* ***Not*** *among Obama's* ***great*** *accomplishments since coming to office !!* [From a tweet in 2010]
  - **Contrastive discourse connectives**: *I used to* ***HATE*** *it.* ***But*** *this stuff is yummmmmy :)* [From a tweet in 2011 -- the tweeter had already bolded “HATE” and “But”!]
- **Multiword expressions**: other words in context can make a negative word positive:
  - That movie was <u>shit</u>. [negative]
  - That movie was <u>the shit</u>. [positive] (American slang from the 1990’s)

More subtle sentiment (from Pang and Lee)

- With many texts, no ostensibly negative words occur, yet they indicate strong negative polarity.
  - *“If you are reading this because it is your darling fragrance, please wear it at home exclusively, and tape the windows shut.*” (review by Luca Turin and Tania Sanchez of the Givenchy perfume Amarige, in *Perfumes: The Guide*, Viking 2008.)
  - “*She runs the gamut of emotions from A to B*.” (Dorothy Parker, speaking about Katharine Hepburn.)
  - *“Jane Austen’s books madden me so that I can’t conceal my frenzy from the reader. Every time I read ‘Pride and Prejudice’ I want to dig her up and beat her over the skull with her own shin-bone*.” (Mark Twain.)

Thwarted expectations (from Pang and Lee)

- There are also highly negative texts that use lots of positive words, but ultimately are reversed by the final sentence. For example

  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.

- This is referred to as a **thwarted expectations narrative** because in the final sentence the author sets up a deliberate contrast to the preceding discourse, giving it more impact.

Thwarted Expectations and Ordering Effects

- "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**."
- Well as usual Keanu Reeves is nothing special, but surprisingly, the very talented Laurence Fishbourne is **not so good** either, I was surprised.

Polarity classification: it’s more than positive and negative

- **Positive**: “As a used vehicle, the Ford Focus represents a solid pick.”
- **Negative**: “Still, the Focus' interior doesn't quite measure up to those offered by some of its competitors, both in terms of materials quality and design aesthetic.”
- **Neutral**: “The Ford Focus has been Ford's entry-level car since the start of the new millennium.”
- **Mixed**: “The current Focus has much to offer in the area of value, if not refinement.”

http://www.edmunds.com/ford/focus/review.html

Other dimensions of sentiment analysis

- **Subjectivity**: is an opinion even being expressed? Many statements are simply factual.
- **Target**: what exactly is an opinion being expressed about?
  - Important for aggregating interesting and meaningful statistics about sentiment.
  - Also, it affects how the language use indicates polarity: e.g, *unpredictable* is usually positive for movie reviews, but is very negative for a car's steering
- **Ratings**: rather than a binary decision, it is often of interest to provide or interpret predictions about sentiment on a scale, such as a 5-star system.

Other dimensions of sentiment analysis

- **Perspective**: an opinion can be positive or negative depending on who is saying it
  - *entry-level* could be good or bad for different people
  - it also affects how an author describes a topic: e.g. *pro-choice* vs *pro-life*, *affordable health care* vs *obamacare*.
- **Authority:** was the text written by someone whose opinion matters more than others?
  - it is more important to identify and address negative sentiment expressed by a popular blogger than a one-off commenter or supplier of a product reviewer on a sales site
  - follower graphs (where applicable) are very useful in this regard
- **Spam**: is the text even valid or at least something of interest?
  - many tweets and blog post comments are just spammers trying to drive traffic to their sites

Sentiment Analysis

A Baseline Algorithm

Sentiment Classification in Movie Reviews

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. EMNLP-2002, 79—86.
Bo Pang and Lillian Lee. 2004. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. ACL, 271-278

- Polarity detection:
  - Is an IMDB movie review positive or negative?
- Data: *Polarity Data 2.0:*
  - http://www.cs.cornell.edu/people/pabo/movie-review-data

IMDB data in the Pang and Lee database

✓

when _star wars_ came out some twenty years ago , the image of traveling throughout the stars has become a commonplace image . […]

when han solo goes light speed , the stars change to bright lines , going towards the viewer in lines that converge at an invisible point .

cool .

october sky offers a much simpler image–that of a single white dot , traveling horizontally across the night sky . [. . .]

✗

" snake eyes " is the most aggravating kind of movie : the kind that shows so much potential then becomes unbelievably disappointing .

it's not just because this is a brian depalma film , and since he's a great director and one who's films are always greeted with at least some fanfare .

and it's not even because this was a film starring nicolas cage and since he gives a brauvara performance , this film is hardly worth his talents .

Baseline Algorithm (adapted from Pang and Lee)

- Tokenization
- Feature Extraction
- Classification using different classifiers
  - Naïve Bayes
  - MaxEnt
  - SVM

Sentiment Tokenization Issues

- Deal with HTML and XML markup
- Twitter mark-up (names, hash tags)
- Capitalization (preserve for words in all caps)
- Phone numbers, dates
- Emoticons
- Useful code:
  - [Christopher Potts sentiment tokenizer]
  - [Brendan O'Connor twitter tokenizer]

Potts emoticons

```
[<>]?                          # optional hat/brow
[:;=8]                         # eyes
[\-o\*\']?                     # optional nose
[\)\]\(\[dDpP/\:\}\{@\|\\]     # mouth
|                              #### reverse orientation
[\)\]\(\[dDpP/\:\}\{@\|\\]     # mouth
[\-o\*\']?                     # optional nose
[:;=8]                         # eyes
[<>]?                          # optional hat/brow
```

Extracting Features for Sentiment Classification

- How to handle negation
  - `I` **`didn't`** `like this movie`

    vs
  - `I really like this movie`
- Which words to use?
  - Only adjectives
  - All words
    - All words turns out to work better, at least on this data

Negation

Das, Sanjiv and Mike Chen. 2001. Yahoo! for Amazon: Extracting market sentiment from stock message boards. In Proceedings of the Asia Pacific Finance Association Annual Conference (APFA).
Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. EMNLP-2002, 79—86.

Add NOT_ to every word between negation and following punctuation:

`didn't like this movie , but I`

`didn't NOT_like NOT_this NOT_movie but I`

Features for classification

That new 300 movie looks sooo friggin BAD ASS .

Features for classification

That new 300 movie looks sooo friggin BAD ASS .

```
w=that
w=new
w=300
w=movie
w=looks
w=sooo
w=friggin
w=bad
w=ass
```

Features for classification

That new 300 movie looks sooo friggin BAD ASS .

```
w=that
w=new
w=300
w=movie
w=looks
w=sooo
w=friggin
w=bad
w=ass
```

```
w=so
```

Features for classification

That new 300 movie looks sooo friggin BAD ASS .

```
w=that
w=new
w=300
w=movie
w=looks
w=sooo
w=friggin
w=bad
w=ass
```

```
w=so
```

```
bi=<START>_that
bi=that_new
bi=new_300
bi=300_movie
bi=movie_looks
bi=looks_sooo
bi=sooo_friggin
bi=friggin_bad
bi=bad_ass
bi=ass_.
bi=._<END>
```

Features for classification

That new 300 movie looks sooo friggin BAD ASS .

art adj noun noun verb adv adv adj noun punc

```
w=that
w=new
w=300
w=movie
w=looks
w=sooo
w=friggin
w=bad
w=ass
```

```
w=so
```

```
bi=<START>_that
bi=that_new
bi=new_300
bi=300_movie
bi=movie_looks
bi=looks_sooo
bi=sooo_friggin
bi=friggin_bad
bi=bad_ass
bi=ass_.
bi=._<END>
```

Features for classification

That new 300 movie looks sooo friggin BAD ASS .

art adj noun noun verb adv adv adj noun punc

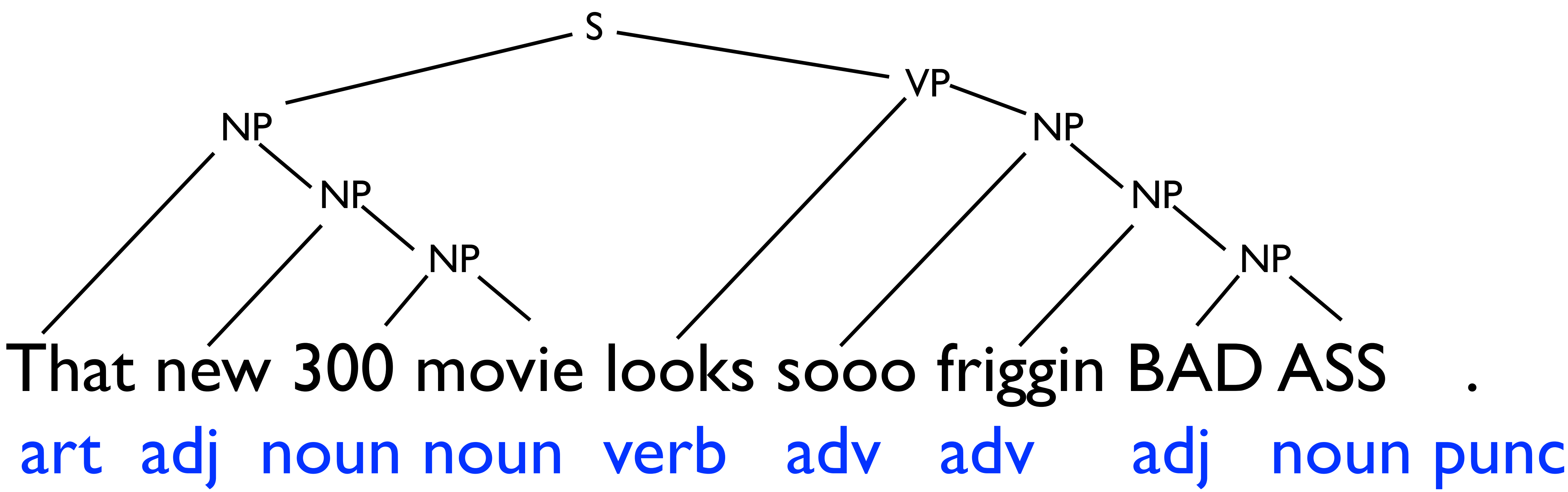
```
w=that
w=new
w=300
w=movie
w=looks
w=sooo
w=friggin
w=bad
w=ass
```

```
w=so
```

```
bi=<START>_that
bi=that_new
bi=new_300
bi=300_movie
bi=movie_looks
bi=looks_sooo
bi=sooo_friggin
bi=friggin_bad
bi=bad_ass
bi=ass_.
bi=._<END>
```

```
wt=that_art
wt=new_adj
wt=300_noun
wt=movie_noun
wt=looks_verb
wt=sooo_adv
wt=friggin_adv
wt=bad_adj
wt=ass_noun
```

Features for classification



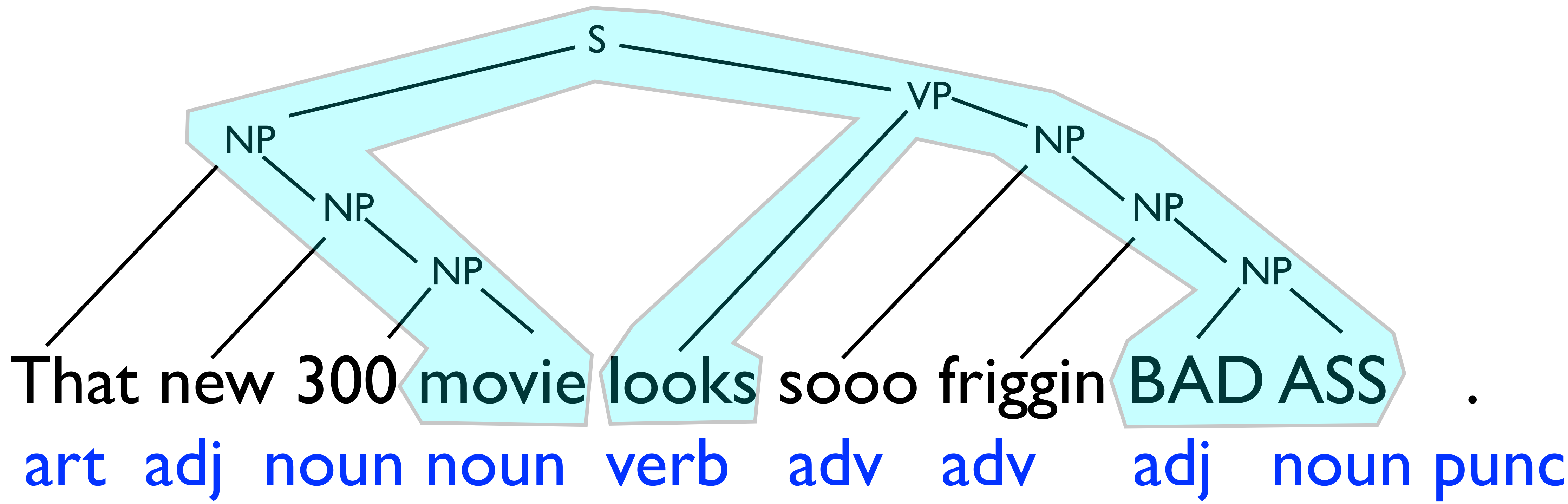

```
w=that
w=new
w=300
w=movie
w=looks
w=sooo
w=friggin
w=bad
w=ass
```

```
w=so
```

```
bi=<START>_that
bi=that_new
bi=new_300
bi=300_movie
bi=movie_looks
bi=looks_sooo
bi=sooo_friggin
bi=friggin_bad
bi=bad_ass
bi=ass_.
bi=._<END>
```

```
wt=that_art
wt=new_adj
wt=300_noun
wt=movie_noun
wt=looks_verb
wt=sooo_adv
wt=friggin_adv
wt=bad_adj
wt=ass_noun
```

Features for classification




```
w=that
w=new
w=300
w=movie
w=looks
w=sooo
w=friggin
w=bad
w=ass
```

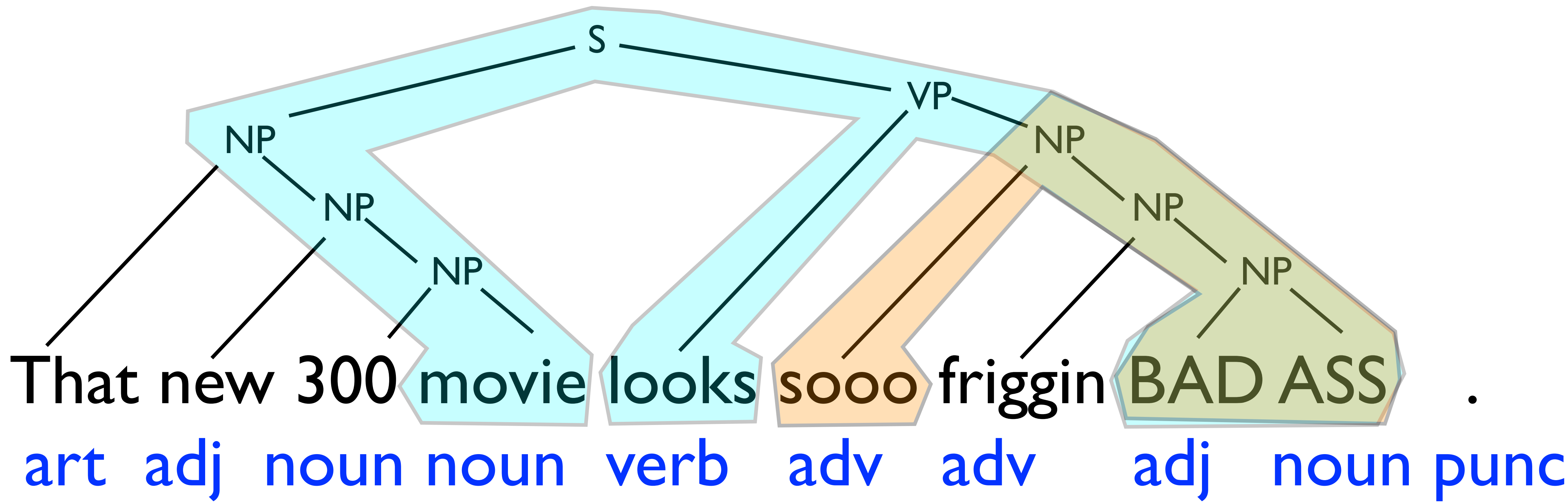
```
w=so
```

```
bi=<START>_that
bi=that_new
bi=new_300
bi=300_movie
bi=movie_looks
bi=looks_sooo
bi=sooo_friggin
bi=friggin_bad
bi=bad_ass
bi=ass_.
bi=._<END>
```

```
wt=that_art
wt=new_adj
wt=300_noun
wt=movie_noun
wt=looks_verb
wt=sooo_adv
wt=friggin_adv
wt=bad_adj
wt=ass_noun
```

```
subtree=S_NP_movie-S_VP_looks-S_VP_NP_bad_ass
```

Features for classification




```
w=that
w=new
w=300
w=movie
w=looks
w=sooo
w=friggin
w=bad
w=ass
```

```
w=so
```

```
bi=<START>_that
bi=that_new
bi=new_300
bi=300_movie
bi=movie_looks
bi=looks_sooo
bi=sooo_friggin
bi=friggin_bad
bi=bad_ass
bi=ass_.
bi=._<END>
```

```
wt=that_art
wt=new_adj
wt=300_noun
wt=movie_noun
wt=looks_verb
wt=sooo_adv
wt=friggin_adv
wt=bad_adj
wt=ass_noun
```

```
subtree=S_NP_movie-S_VP_looks-S_VP_NP_bad_ass
```

```
subtree=NP_sooo_bad_ass
```

Complexity of features

- Features can be defined on very deep aspects of the linguistic content, including syntactic and rhetorical structure.
- The models for these can be quite complex, and often require significant training material to learn them, which means it is harder to employ them for languages without such resources.
  - I'll show an example for part-of-speech tagging in a bit.
- Also: the more fine-grained the feature, the more likely it is rare to see in one's training corpus. This requires more training data, or effective semi-supervised learning methods.

Sentiment Analysis

Sentiment Lexicons

The General Inquirer

Philip J. Stone, Dexter C Dunphy, Marshall S. Smith, Daniel M. Ogilvie. 1966. The General Inquirer: A Computer Approach to Content Analysis. MIT Press

- Home page: http://www.wjh.harvard.edu/~inquirer
- List of Categories: http://www.wjh.harvard.edu/~inquirer/homecat.htm
- Spreadsheet: http://www.wjh.harvard.edu/~inquirer/inquirerbasic.xls
- Categories:
  - Positiv (1915 words) and Negativ (2291 words)
  - Strong vs Weak, Active vs Passive, Overstated versus Understated
  - Pleasure, Pain, Virtue, Vice, Motivation, Cognitive Orientation, etc
- Free for Research Use

LIWC (Linguistic Inquiry and Word Count)

Pennebaker, J.W., Booth, R.J., & Francis, M.E. (2007). Linguistic Inquiry and Word Count: LIWC 2007. Austin, TX

- Home page: http://www.liwc.net/
- 2300 words, >70 classes
- **Affective Processes**
  - negative emotion (*bad, weird, hate, problem, tough*)
  - positive emotion (*love, nice, sweet*)
- **Cognitive Processes**
  - Tentative (*maybe, perhaps, guess*), Inhibition (*block, constraint*)
- **Pronouns, Negation** (*no, never*), **Quantifiers** (*few, many*)
- $30 or $90 fee

MPQA Subjectivity Cues Lexicon

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. Proc. of HLT-EMNLP-2005.

Riloff and Wiebe (2003). Learning extraction patterns for subjective expressions. EMNLP-2003.

- Home page: http://www.cs.pitt.edu/mpqa/subj_lexicon.html
- 6885 words from 8221 lemmas
  - 2718 positive
  - 4912 negative
- Each word annotated for intensity (strong, weak)
- GNU GPL

Bing Liu Opinion Lexicon

Minqing Hu and Bing Liu. Mining and Summarizing Customer Reviews. ACM SIGKDD-2004.

- Bing Liu's Page on Opinion Mining
- http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar

- 6786 words
  - 2006 positive
  - 4783 negative

SentiWordNet

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010 SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. LREC-2010

- Home page: http://sentiwordnet.isti.cnr.it/
- All WordNet synsets automatically annotated for degrees of positivity, negativity, and neutrality/objectiveness
- [estimable(J,3)] "may be computed or estimated"

```
Pos  0   Neg 0   Obj 1
```

- [estimable(J,1)] "deserving of respect or high regard"

```
Pos .75  Neg 0   Obj .25
```

Disagreements between polarity lexicons

Christopher Potts, Sentiment Tutorial, 2011

	Opinion Lexicon	General 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%)**
General Inquirer			520/2306 **(23%)**	1/204 **(0.5%)**
SentiWordNet				174/694 **(25%)**
LIWC				

Dan Jurafsky

Analyzing the polarity of each word in IMDB

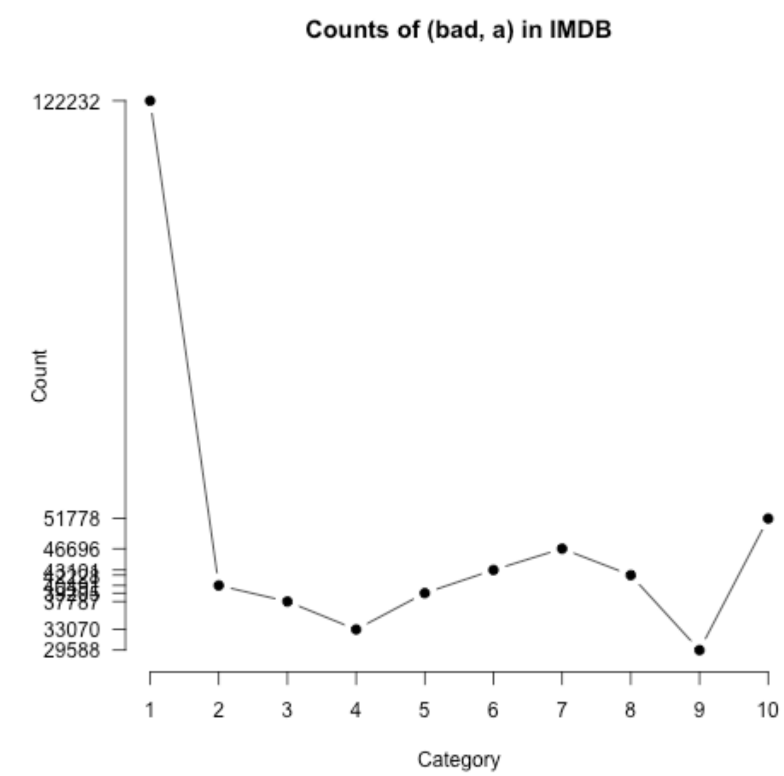
Potts, Christopher. 2011. On the negativity of negation. SALT 20, 636-659.

- How likely is each word to appear in each sentiment class?
- Count("bad") in 1-star, 2-star, 3-star, etc.
- But can't use raw counts:
- Instead, **likelihood:**

$$P(w \mid c) = \frac{f(w,c)}{\sum_{w \in c} f(w,c)}$$

- Make them comparable between words
  - **Scaled likelihood:**

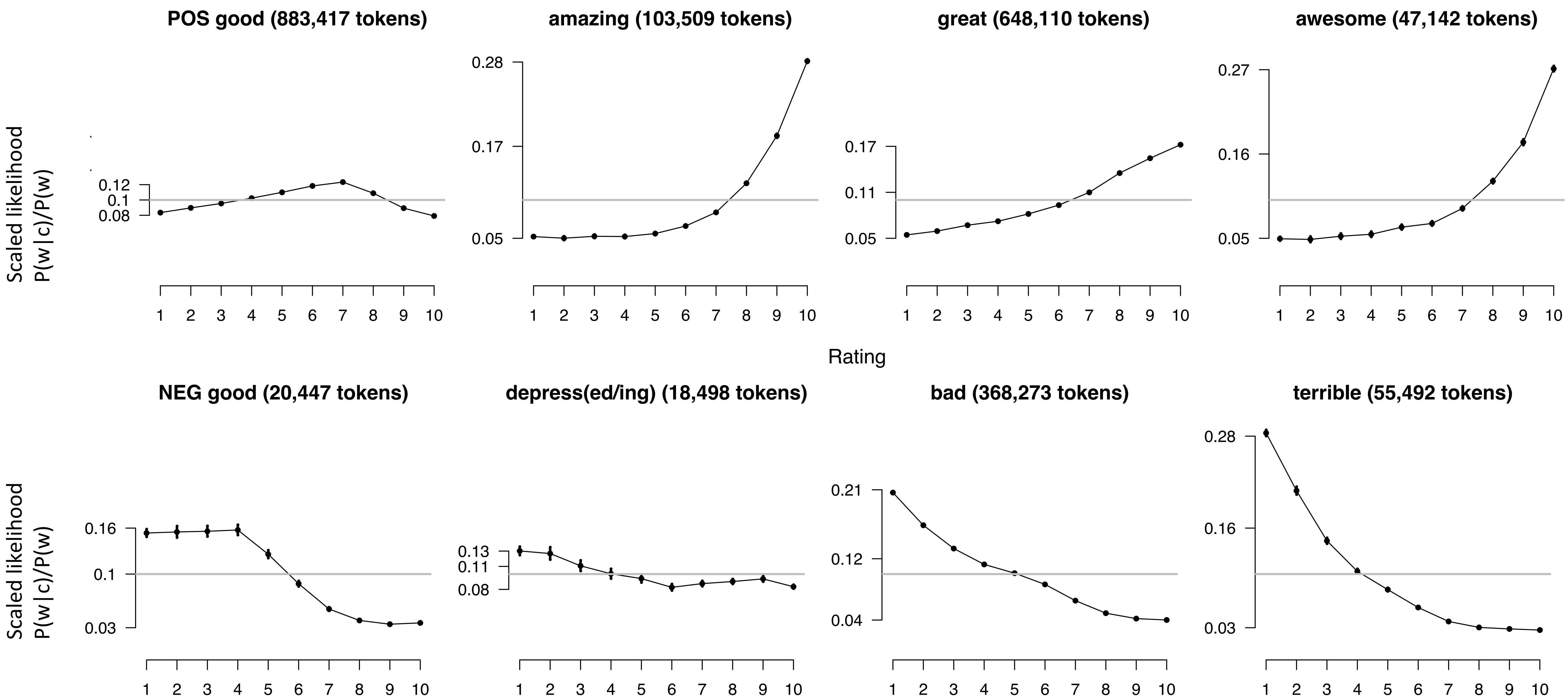
$$\frac{P(w \mid c)}{P(w)}$$



Dan Jurafsky

Analyzing the polarity of each word in IMDB

Potts, Christopher. 2011. On the negativity of negation. SALT 20, 636-659.



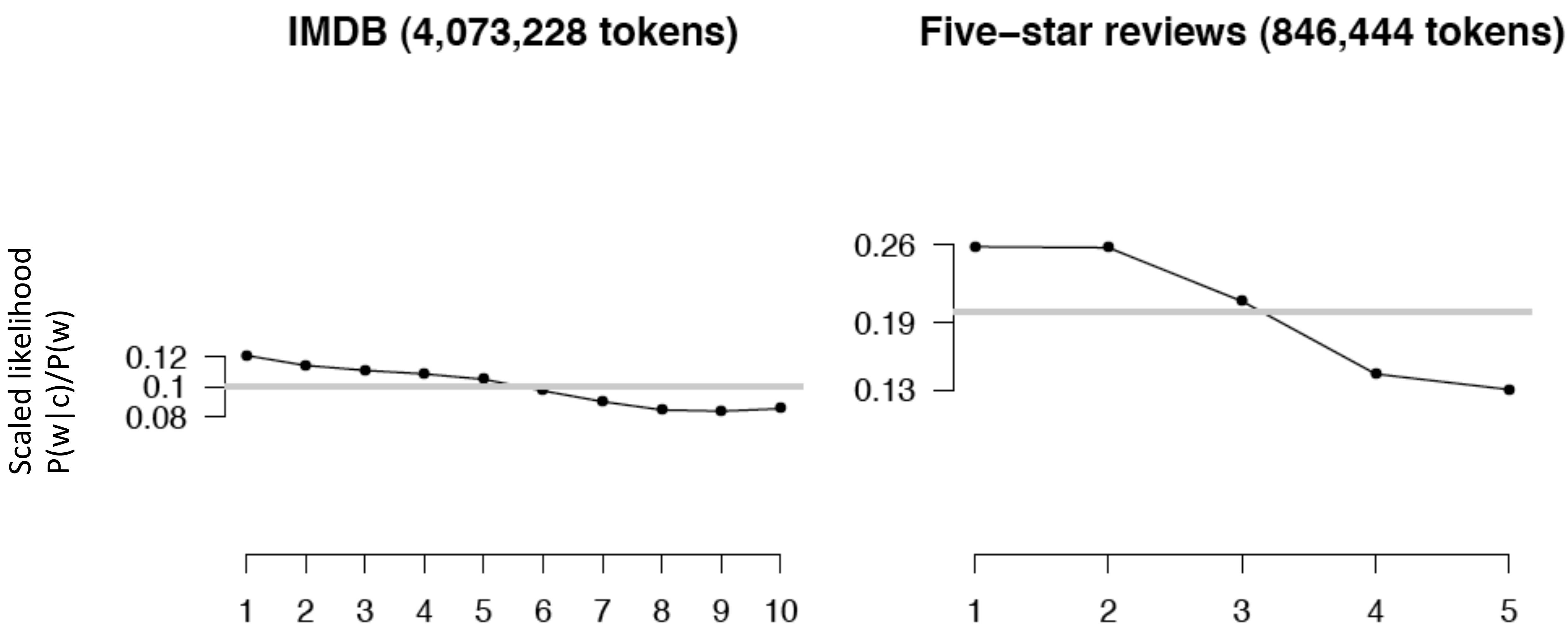
Other sentiment feature: Logical negation

Potts, Christopher. 2011. On the negativity of negation. SALT 20, 636-659.

- Is logical negation (*no, not*) associated with negative sentiment?
- Potts experiment:
  - Count negation (*not, n't, no, never*) in online reviews
  - Regress against the review rating

Dan Jurafsky

Potts 2011 Results: More negation in negative sentiment



Sentiment Analysis

Learning Sentiment Lexicons

Semi-supervised learning of lexicons

- Use a small amount of information
  - A few labeled examples
  - A few hand-built patterns
- To bootstrap a lexicon

Hatzivassiloglou and McKeown intuition for identifying word polarity

Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the Semantic Orientation of Adjectives. ACL, 174–181

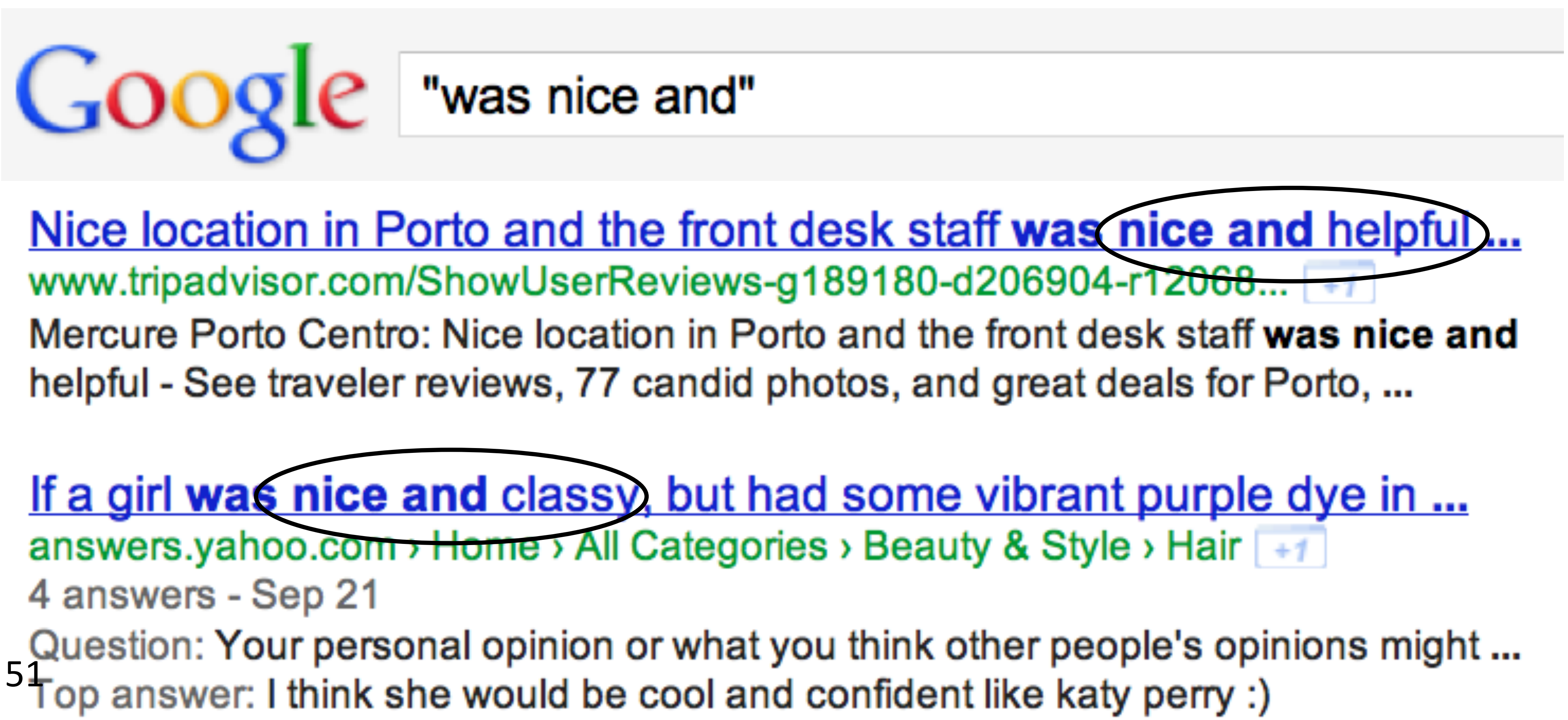
- Adjectives conjoined by "*and*" have same polarity
  - Fair **and** legitimate, corrupt **and** brutal
  - *fair **and** brutal, *corrupt **and** legitimate
- Adjectives conjoined by "*but*" do not
  - fair **but** brutal

Hatzivassiloglou & McKeown 1997
Step 1

- Label **seed set** of 1336 adjectives (all >20 in 21 million word WSJ corpus)
  - 657 positive
    - adequate central clever famous intelligent remarkable reputed sensitive slender thriving…
  - 679 negative
    - contagious drunken ignorant lanky listless primitive strident troublesome unresolved unsuspecting…

Hatzivassiloglou & McKeown 1997 Step 2

- Expand seed set to conjoined adjectives

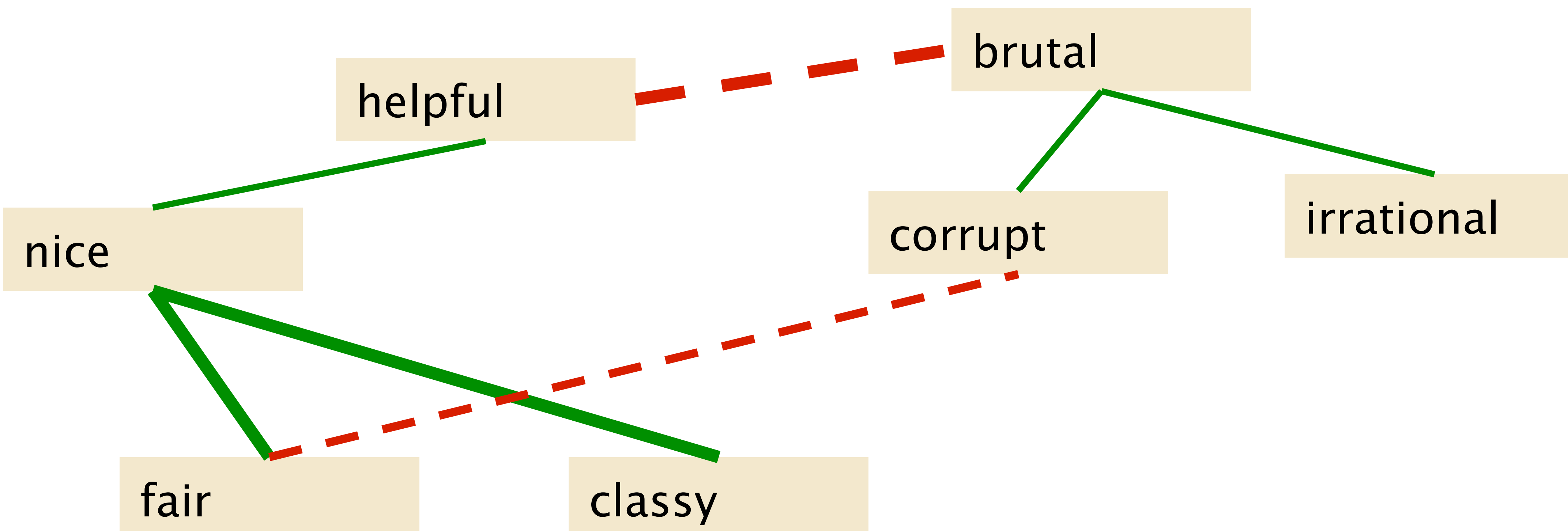


nice, helpful

nice, classy

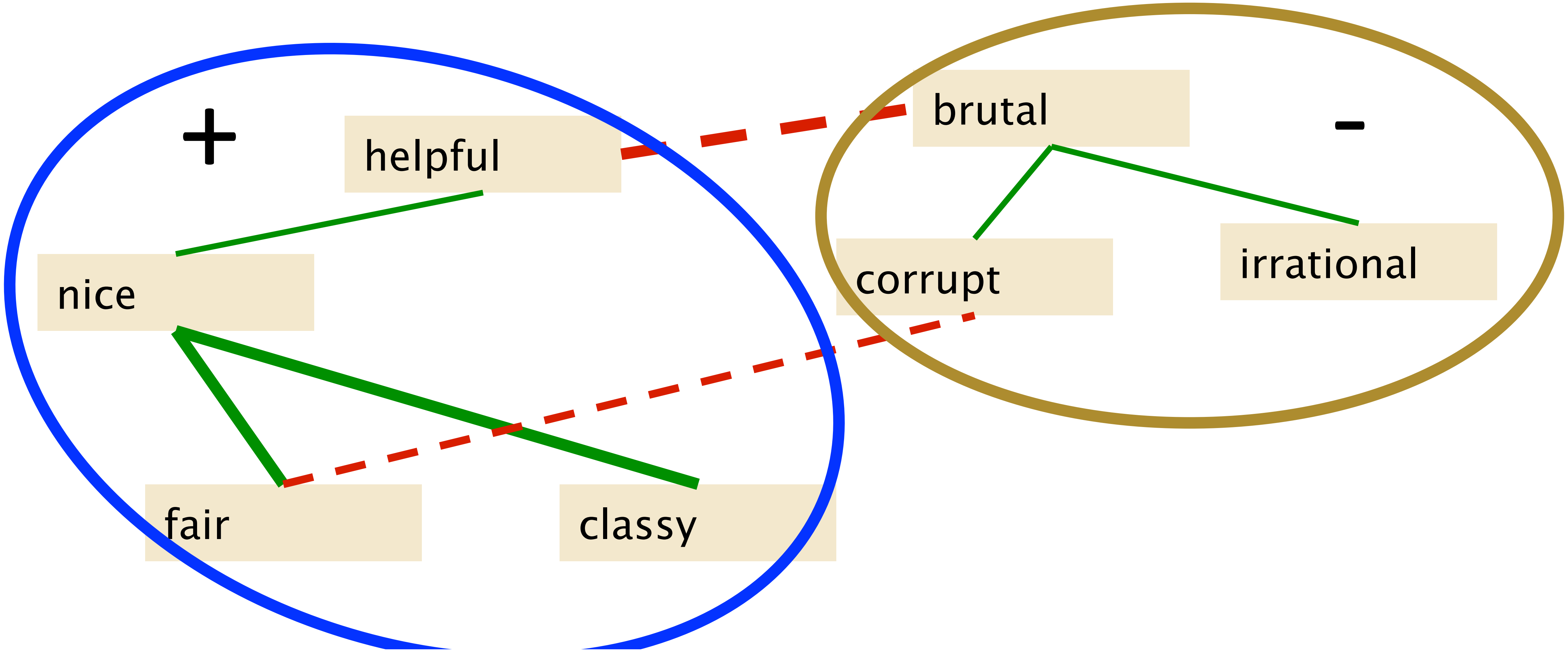
Hatzivassiloglou & McKeown 1997
Step 3

- Supervised classifier assigns "polarity similarity" to each word pair, resulting in graph:



Hatzivassiloglou & McKeown 1997
Step 4

- Clustering for partitioning the graph into two



Dan Jurafsky

Output polarity lexicon

- Positive
  - bold decisive disturbing generous good honest important large mature patient peaceful positive proud sound stimulating straightforward strange talented vigorous witty…
- Negative
  - ambiguous cautious cynical evasive harmful hypocritical inefficient insecure irrational irresponsible minor outspoken pleasant reckless risky selfish tedious unsupported vulnerable wasteful…

Output polarity lexicon

- Positive
  - bold decisive **disturbing** generous good honest important large mature patient peaceful positive proud sound stimulating straightforward **strange** talented vigorous witty…
- Negative
  - ambiguous **cautious** cynical evasive harmful hypocritical inefficient insecure irrational irresponsible minor **outspoken pleasant** reckless risky selfish tedious unsupported vulnerable wasteful…

Turney Algorithm

Turney (2002): Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews

1. Extract a *phrasal lexicon* from reviews
2. Learn polarity of each phrase
3. Rate a review by the average polarity of its phrases

Dan Jurafsky

Extract two-word phrases with adjectives

First Word	Second Word	Third Word (not extracted)
JJ	NN or NNS	anything
RB, RBR, RBS	JJ	Not NN nor NNS
JJ	JJ	Not NN or NNS
NN or NNS	JJ	Nor NN nor NNS
RB, RBR, or RBS	VB, VBD, VBN, VBG	anything

How to measure polarity of a phrase?

- Positive phrases co-occur more with *"excellent"*
- Negative phrases co-occur more with *"poor"*
- But how to measure co-occurrence?

Pointwise Mutual Information

- **Mutual information** between 2 random variables X and Y

$$\mathrm{I}(X,Y) = \sum_x \sum_y P(x,y) \log_2 \frac{P(x,y)}{P(x)P(y)}$$

- **Pointwise mutual information**:
  - How much more do events x and y co-occur than if they were independent?

$$\mathrm{PMI}(X,Y) = \log_2 \frac{P(x,y)}{P(x)P(y)}$$

Pointwise Mutual Information

- **Pointwise mutual information**:
  - How much more do events x and y co-occur than if they were independent?

$$\mathrm{PMI}(X,Y) = \log_2 \frac{P(x,y)}{P(x)P(y)}$$

- **PMI between two words**:
  - How much more do two words co-occur than if they were independent?

$$\mathrm{PMI}(word_1, word_2) = \log_2 \frac{P(word_1, word_2)}{P(word_1)P(word_2)}$$

Dan Jurafsky

How to Estimate Pointwise Mutual Information

- Query search engine (Altavista)
  - P(word) estimated by `hits(word)/N`
  - $P(word_1,word_2)$ by `hits(word1 NEAR word2)/N`
    - (More correctly the bigram denominator should be kN, because there are a total of N consecutive bigrams (word1,word2), but kN bigrams that are k words apart, but we just use N on the rest of this slide and the next.)

$$\mathrm{PMI}(word_1, word_2) = \log_2 \frac{\frac{1}{N} hits(word_1 \text{ NEAR } word_2)}{\frac{1}{N} hits(word_1) \frac{1}{N} hits(word_2)}$$

Does phrase appear more with "poor" or "excellent"?

$$\text{Polarity}(\textit{phrase}) = \text{PMI}(\textit{phrase}, \text{"excellent"}) - \text{PMI}(\textit{phrase}, \text{"poor"})$$

$$= \log_2 \frac{\frac{1}{N}\text{hits}(\textit{phrase} \text{ NEAR "excellent"})}{\frac{1}{N}\text{hits}(\textit{phrase})\frac{1}{N}\text{hits}(\text{"excellent"})} - \log_2 \frac{\frac{1}{N}\text{hits}(\textit{phrase} \text{ NEAR "poor"})}{\frac{1}{N}\text{hits}(\textit{phrase})\frac{1}{N}\text{hits}(\text{"poor"})}$$

$$= \log_2 \frac{\text{hits}(\textit{phrase} \text{ NEAR "excellent"})}{\text{hits}(\textit{phrase})\text{hits}(\text{"excellent"})} \frac{\text{hits}(\textit{phrase})\text{hits}(\text{"poor"})}{\text{hits}(\textit{phrase} \text{ NEAR "poor"})}$$

$$= \log_2 \left(\frac{\text{hits}(\textit{phrase} \text{ NEAR "excellent"})\text{hits}(\text{"poor"})}{\text{hits}(\textit{phrase} \text{ NEAR "poor"})\text{hits}(\text{"excellent"})} \right)$$

Dan Jurafsky

Phrases from a thumbs-up review

Phrase	POS tags	Polarity
online service	JJ NN	2.8
online experience	JJ NN	2.3
direct deposit	JJ NN	1.3
local branch	JJ NN	0.42
…		
low fees	JJ NNS	0.33
true service	JJ NN	-0.73
other bank	JJ NN	-0.85
inconveniently located	JJ NN	-1.5
Average		0.32

Phrases from a thumbs-down review

Phrase	POS tags	Polarity
direct deposits	JJ NNS	5.8
online web	JJ NN	1.9
very handy	RB JJ	1.4
…		
virtual monopoly	JJ NN	-2.0
lesser evil	RBR JJ	-2.3
other problems	JJ NNS	-2.8
low funds	JJ NNS	-6.8
unethical practices	JJ NNS	-8.5
Average		-1.2

Results of Turney algorithm

- 410 reviews from Epinions
  - 170 (41%) negative
  - 240 (59%) positive
- Majority class baseline: 59%
- Turney algorithm: 74%

- Phrases rather than words
- Learns domain-specific information

Using WordNet to learn polarity

S.M. Kim and E. Hovy. 2004. Determining the sentiment of opinions. COLING 2004
M. Hu and B. Liu. Mining and summarizing customer reviews. In Proceedings of KDD, 2004

- WordNet: online thesaurus (covered in later lecture).
- Create positive ("good") and negative seed-words ("terrible")
- Find Synonyms and Antonyms
  - Positive Set: Add synonyms of positive words ("well") and antonyms of negative words
  - Negative Set: Add synonyms of negative words ("awful") and antonyms of positive words ("evil")
- Repeat, following chains of synonyms
- Filter

Dan Jurafsky

Summary on Learning Lexicons

- Advantages:
  - Can be domain-specific
  - Can be more robust (more words)
- Intuition
  - Start with a seed set of words ('good', 'poor')
  - Find other words that have similar polarity:
    - Using "and" and "but"
    - Using words that occur nearby in the same document
    - Using WordNet synonyms and antonyms

Sentiment Analysis

Other Sentiment Tasks

Finding sentiment of a sentence

- Important for finding aspects or attributes
  - Target of sentiment

- `The food was great but the service was awful`

Finding aspect/attribute/target of sentiment

M. Hu and B. Liu. 2004. Mining and summarizing customer reviews. In Proceedings of KDD.
S. Blair-Goldensohn, K. Hannan, R. McDonald, T. Neylon, G. Reis, and J. Reynar. 2008. Building a Sentiment Summarizer for Local Service Reviews. WWW Workshop.

- Frequent phrases + rules
  - Find all highly frequent phrases across reviews ("`fish tacos`")
  - Filter by rules like "occurs right after sentiment word"
    - "…`great fish tacos`" means `fish tacos` a likely aspect

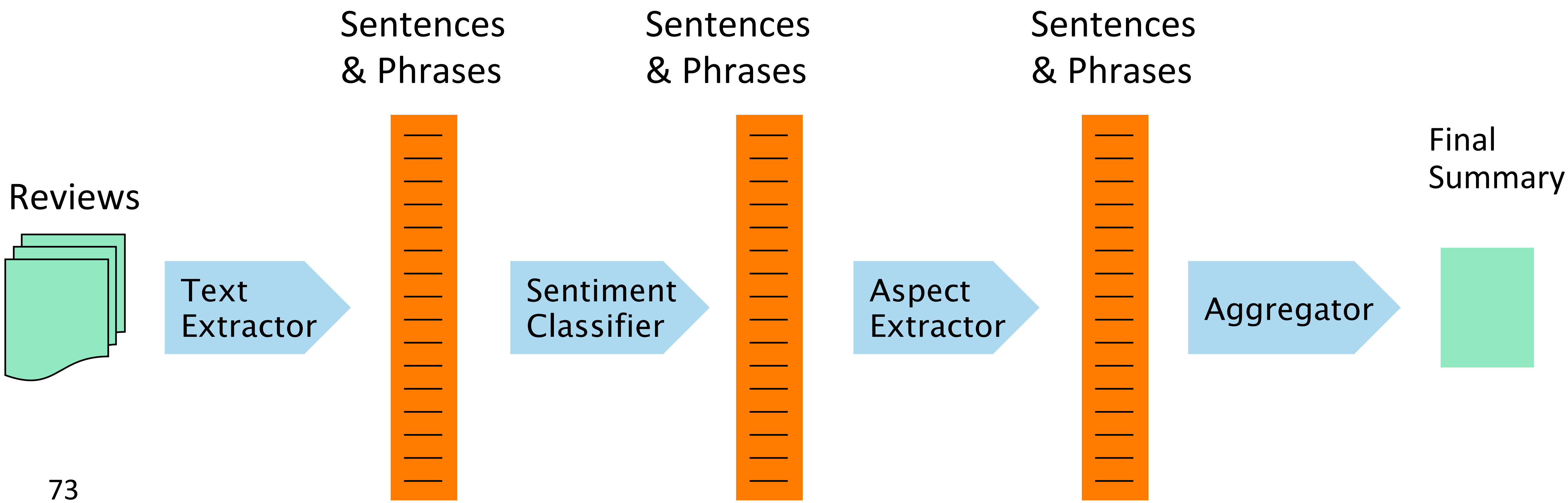
Casino	casino, buffet, pool, resort, beds
Children's Barber	haircut, job, experience, kids
Greek Restaurant	food, wine, service, appetizer, lamb
Department Store	selection, department, sales, shop, clothing

Finding aspect/attribute/target of sentiment

- The aspect name may not be in the sentence
- For restaurants/hotels, aspects are well-understood
- Supervised classification
  - Hand-label a small corpus of restaurant review sentences with aspect
    - food, décor, service, value, NONE
  - Train a classifier to assign an aspect to asentence
    - "Given this sentence, is the aspect *food, décor, service, value,* or *NONE*"

Putting it all together: Finding sentiment for aspects

S. Blair-Goldensohn, K. Hannan, R. McDonald, T. Neylon, G. Reis, and J. Reynar. 2008. Building a Sentiment Summarizer for Local Service Reviews. WWW Workshop



Results of Blair-Goldensohn et al. method

Rooms (3/5 stars, 41 comments)

- (+) The room was clean and everything worked fine – even the water pressure ...
- (+) We went because of the free room and was pleasantly pleased ...
- (-) ...the worst hotel I had ever stayed at ...

Service (3/5 stars, 31 comments)

- (+) Upon checking out another couple was checking early due to a problem ...
- (+) Every single hotel staff member treated us great and answered every ...
- (-) The food is cold and the service gives new meaning to SLOW.

Dining (3/5 stars, 18 comments)

- (+) our favorite place to stay in biloxi.the food is great also the service ...
- (+) Offer of free buffet for joining the Play

Baseline methods assume classes have equal frequencies!

- If not balanced (common in the real world)
  - can't use accuracies as an evaluation
  - need to use F-scores
- Severe imbalancing also can degrade classifier performance
- Two common solutions:
  1. Resampling in training
     - Random undersampling
  2. Cost-sensitive learning
     - Penalize SVM more for misclassification of the rare thing

How to deal with 7 stars?

Bo Pang and Lillian Lee. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. ACL, 115–124

1. Map to binary
2. Use linear or ordinal regression
   - Or specialized models like metric labeling

Dan Jurafsky

Summary on Sentiment

- Generally modeled as classification or regression task
  - predict a binary or ordinal label
- Features:
  - Negation is important
  - Using all words (in naïve bayes) works well for some tasks
  - Finding subsets of words may help in other tasks
    - Hand-built polarity lexicons
    - Use seeds and semi-supervised learning to induce lexicons

Scherer Typology of Affective States

- **Emotion**: brief organically synchronized … evaluation of a major event
  - *angry, sad, joyful, fearful, ashamed, proud, elated*
- **Mood**: diffuse non-caused low-intensity long-duration change in subjective feeling
  - *cheerful, gloomy, irritable, listless, depressed, buoyant*
- **Interpersonal stances**: affective stance toward another person in a specific interaction
  - *friendly, flirtatious, distant, cold, warm, supportive, contemptuous*
- **Attitudes**: enduring, affectively colored beliefs, dispositions towards objects or persons
  - *liking, loving, hating, valuing, desiring*
- **Personality traits**: stable personality dispositions and typical behavior tendencies
  - *nervous, anxious, reckless, morose, hostile, jealous*

Computational work on other affective states

- **Emotion**:
  - Detecting annoyed callers to dialogue system
  - Detecting confused/frustrated versus confident students
- **Mood**:
  - Finding traumatized or depressed writers
- **Interpersonal stances**:
  - Detection of flirtation or friendliness in conversations
- **Personality traits**:
  - Detection of extroverts

Detection of Friendliness

Ranganath, Jurafsky, McFarland

- Friendly speakers use collaborative conversational style
  - Laughter
  - Less use of negative emotional words
  - More sympathy
    - `That's too bad` `I'm sorry to hear that`
  - More agreement
    - `I think so too`
  - Less hedges
    - `kind of` `sort of` `a little …`

how to make a racist AI?

(notebook)

Learning about the World / biases in text

(slides)

Ethics

think about what you do.

Ethics

- Who is going to use your system and why?
- Who may get harmed from your system?
    - Intentionally
    - Unintentionally
- Who will be excluded?

Ethics

- What if your system is 100% accurate?
- What if its 90% accurate?
- How are the mistakes distributed?

Ethics

- What biases do you encode into your system?