Sentiment Analysis

Dr. Demetrios Glinos
University of Central Florida

CAP6640 – Computer Understanding of Natural Language

Today

- What is Sentiment Analysis?
- A Baseline Algorithm
- Sentiment Lexicons
- Learning Sentiment Lexicons
- Other Sentiment Tasks

Movie reviews

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

- Unbelievably disappointing

- This is the greatest screwball comedy ever filmed

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

Google Shopping

Epson WorkForce WF-2760 All-In-One Color Inkjet Printer - Wireless - Black My Shortlist (0)

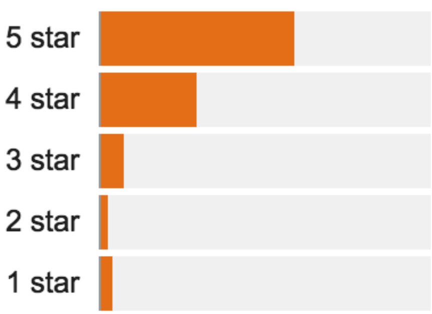
$70 online, **$70** nearby

#2 in Epson Printers, Copiers & Fax Machines ★★★★★ 1,559 product reviews Save to Shortlist

Epson · WorkForce Series · WorkForce · All-in-One · Inkjet · Color · Wireless · With Scanner · 13 ppm (mono) · Wi-Fi

This wireless all-in-one unit can work with documents and paper up to 8.5 x 47.2 and produces highly detailed prints with a maximum resolution of 4800 x 1200 dpi. It has a print speed of 13.7 ppm in black or 7.3 ppm in color and can create borderless photos at sizes up to 8.5 x 11. Automatic duplex printing is also available. Scanning is covered with a 2400 dpi.
« less

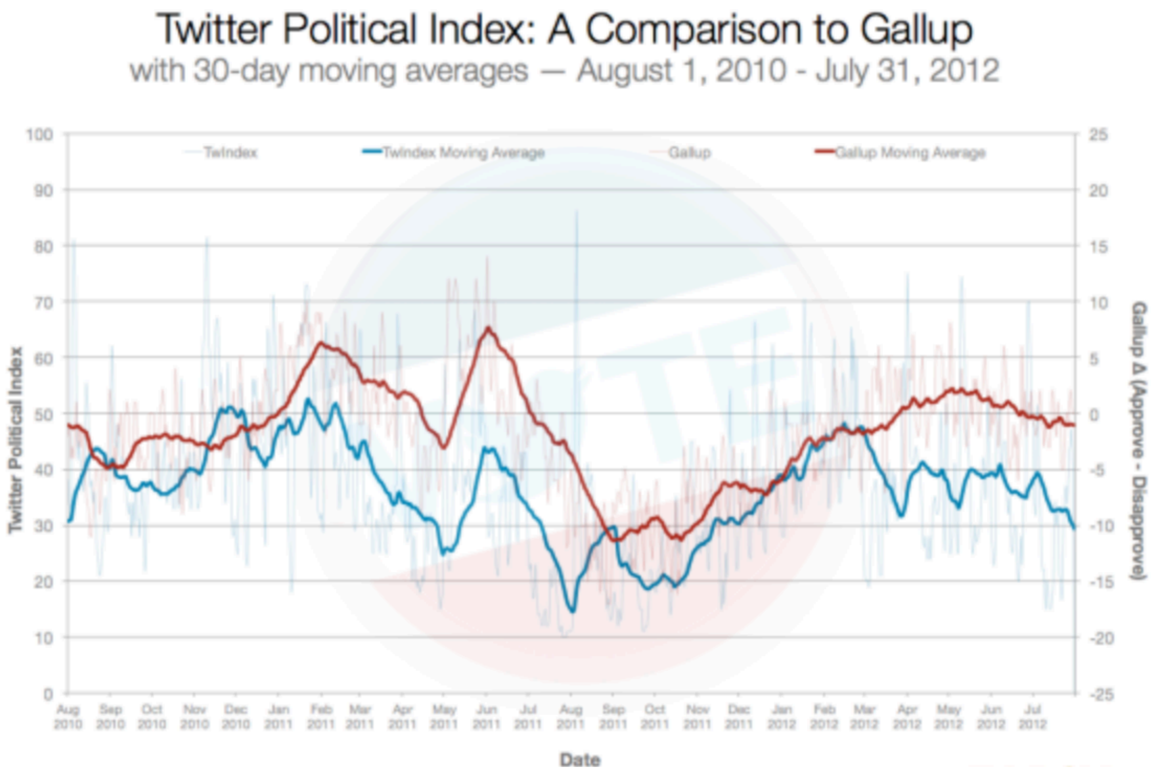
Reviews



4.4
★★★★★
1,559 reviews

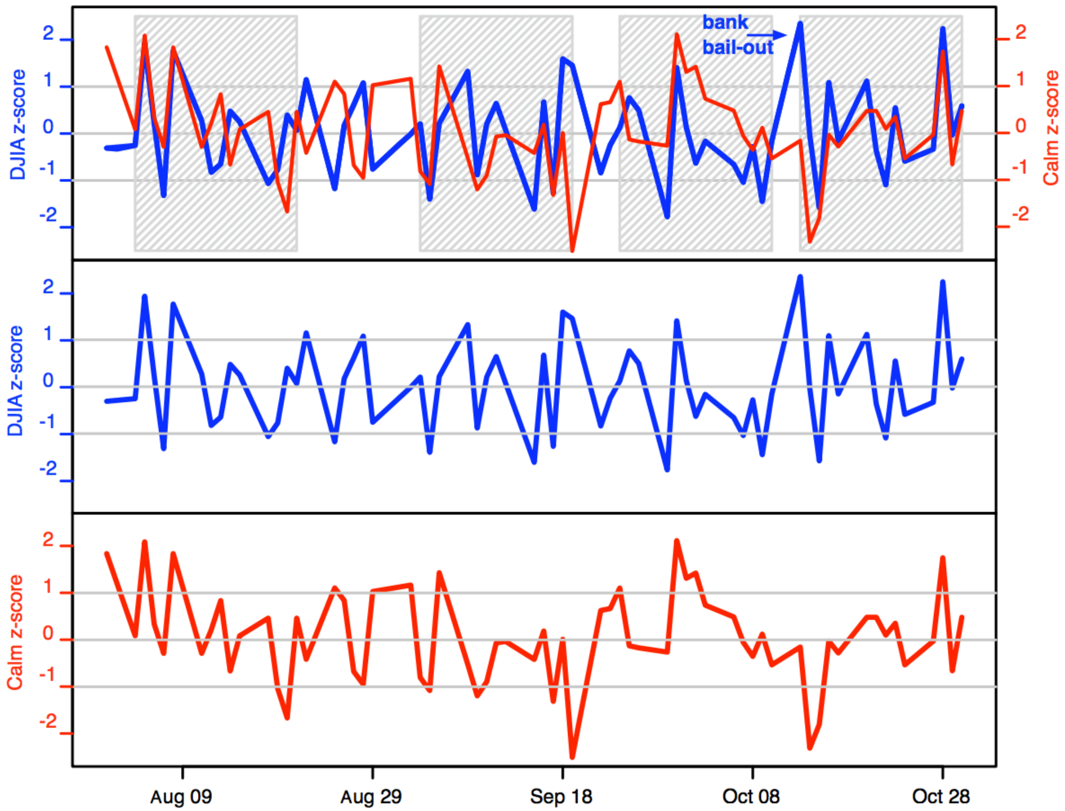
Ease of use	"Overall easy to install and great quality printing."
Value	"Great price, sleek design."
Setup	"Easy setup easy operation."
Picture/video	"Great resolution and quality!"
Size	"Nice size not too big."
Colors	"I love the ink colors."
Design/style	"The design is compact and attractive."

Twitter Sentiment v. Gallup Poll



- source: https://www.theverge.com/2012/8/1/3213129/twitter-political-index-romney-obama-tweet-comparison
- see also: "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series", B. O'Connor, R. Balasubramanyan, B. Routledge, N. Smith (ICWSM 2010)

Twitter Sentiment v. Stock Market



source: "Twitter mood predicts the stock market", J. Bollen, H. Mao, X. Zeng (J. Computational Science, 2011)

Note: Mood states examined: calm, alert, sure, vital, kind, happy

Why do sentiment analysis?

- Movie reviews
  - discrete or continuous scales
- Product reviews
  - overall, plus feature-based
- Politics
  - determining what people think
  - can be time-sensitive
- Prediction
  - predicting election outcomes, market trends, etc.

Other names for sentiment analysis

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

Scherer Typology of Affective States

Emotion: Relatively brief episode of response to the evaluation of an external or internal event as being of major significance.
(*angry, sad, joyful, fearful, ashamed, proud, elated, desperate*)

Mood: Diffuse affect state, most pronounced as change in subjective feeling, of low intensity but relatively long duration, often without apparent cause.
(*cheerful, gloomy, irritable, listless, depressed, buoyant*)

Interpersonal stance: Affective stance taken toward another person in a specific interaction, colouring the interpersonal exchange in that situation.
(*distant, cold, warm, supportive, contemptuous, friendly*)

Attitude: Relatively enduring, affectively colored beliefs, preferences, and predispositions towards objects or persons.
(*liking, loving, hating, valuing, desiring*)

Personality traits: Emotionally laden, stable personality dispositions and behavior tendencies, typical for a person.
(*nervous, anxious, reckless, morose, hostile, jealous*)

source: Fig. 18.1, Speech and Language Processing. Daniel Jurafsky & James H. Martin. (3rd Ed. Draft November 7, 2016)

Sentiment Analysis

- Detection of **attitude**
  - "relatively enduring, affectively colored beliefs, preferences, and predispositions towards objects or persons"

- an attitude is characterized by
  - Holder (source) – who has the attitude
  - Target (aspect) – about what or whom
  - Type can be
    - from a set of types: like, love, hate, value, desire
    - or, more commonly, a simple weighted polarity
      - positive, negative, neutral, together with strength
  - Text
    - expresses the attitude
    - can be a sentence or an entire document

Sentiment Analysis Tasks

- Simple
  - 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 type

Today

- What is Sentiment Analysis?
- A Baseline Algorithm
- Sentiment Lexicons
- Learning Sentiment Lexicons
- Other Sentiment Tasks

Sentiment Classification in Movie Reviews

Thumbs up? Sentiment Classification using Machine Learning Techniques

Bo Pang and **Lillian Lee**
Department of Computer Science
Cornell University
Ithaca, NY 14853 USA
{pabo,llee}@cs.cornell.edu

Shivakumar Vaithyanathan
IBM Almaden Research Center
650 Harry Rd.
San Jose, CA 95120 USA
shiv@almaden.ibm.com

source: Proceedings EMNLP 2002

- Polarity detection
  - Is an IMDb movie review positive or negative?

- Data set
  - Polarity Data 2.0
  - http://www.cs.cornell.edu/people/pabo/movie-review-data/

IMDb excerpts from Polarity 2.0 data set

- Positive review
  - "every now and then a movie comes along from a suspect studio , with every indication that it will be a stinker , and to everybody's surprise (perhaps even the studio) the film becomes a critical darling . "
- Negative review
  - "this is a by-the-books movie that plods along on a predestined course with no surprises and very few laughs . it also jumps on the ever-popular political satire bandwagon and manages to fall flat there , too . ..."

Note: neither of these excerpts uses obvious keywords like "great", "like", hate", etc.

Baseline Algorithm (adapted from Pang and Lee)

- Tokenization
- Feature extraction
- Classification using different classifiers
  - Naïve Bayes
  - MaxEnt
  - SVM
  - *Note: Pang and Lee tried all three classifiers and found all three did better than human-crafted rules, but they all performed worse on sentiment analysis than on other NLP tasks*

Tokenization Issues

- Handling HTML and XML markup
- Twitter markup (names, hash tags)
- Capitalization
  - should preserve for words in all caps
- Handling phone numbers and dates
- What to do with emoticons

- Useful code for Twitter:
  - http://sentiment.christopherpotts.net/code-data/happyfuntokenizing.py

- See also http://www.cs.cmu.edu/~ark/TweetNLP/ for a suite containing tokenizer, tagger, and dependency parser for tweets, along with annotated corpora

Issues in Feature Extraction

- Handling negation

```
I didn't like this movie
I really like this movie
```

- Which words to include in the analysis
  - only adjectives (e.g., "a **good** movie")
  - all words
    - all words works better (at least on movie reviews)

Handling Negation

- Handling negation is a problem for text understanding generally
  - "not", "never", "no"

- For short texts, surface processing can be helpful
  - add marker to tokens following the detected negation, up to next punctuation

  - Example:

```
I didn't like this movie , but I
I didn't NOT_like NOT_this NOT_movie , but I
```

see: "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web," Sanjiv Das and Mike Chen (2004)
https://pdfs.semanticscholar.org/89af/734dad7645ca3f3a7717443a1e49f37964bc.pdf

Recall: Multinomial Naïve Bayes

$$c_{NB} = argmax_{c_j \in C}\, P(c_j) \prod_{i=1}^{n} P(\, w_i \mid c_j \,)$$

where:

$$P(\, c_j \,) = \frac{|\, docsj \,|}{|\, total\ \#\ documents \,|}$$

$$\hat{P}(\, w_k \mid c_j) = \frac{count\big(w_k, c_j\,\big) + 1}{(\sum_{w \in V} count(\, w, c_j\,)) + |V|}$$

Binarized (Boolean) Multinomial Naïve Bayes

- Word *occurrence* may matter more than word *frequency*
  - Example:
    - "fantastic" occurs in a review → this tells us a great deal
    - if it occurs 5 times in the review, this does not tell us much more

- Solution: Boolean Multinomial Naïve Bayes
  - clip all word counts in each document at 1
  - this seems to work better for sentiment analysis than full word counts

Boolean Naïve Bayes Learning

- From the training corpus, extract the *Vocabulary*

- Calculate $P(c_j)$ terms

  for each c_j in C, find $docs_j$ = { all docs with class = c_j }

$$P(c_j) = \frac{|\,docsj\,|}{|\,total\ \#\ documents\,|}$$

- Calculate $P(w_k \mid c_j)$ terms

  Remove duplicate words in all documents

  Let $Text_j$ = a mega-doc containing all docs in doc_j

  for each word w_k in *Vocabulary*, compute

  n_k = # occurrences of w_k in $Test_j$

$$P(wk \mid c_j) = \frac{n_k + \alpha}{n + \alpha|Vocabulary|}$$

where α = 1 for Laplace

Boolean Naïve Bayes Testing

- Given a test document **d**
  - First remove all duplicate words from **d**
  - Then compute Naïve Bayes using the same equation as before

$$c_{NB} = argmax_{c_j \in C}\, P(c_j) \prod_{i=1}^{n} P(\, w_i \mid c_j \,)$$

Normal v. Boolean Multinomial Naïve Bayes

Source documents:

(0) [+] I liked the movie
(1) [−] I hated the movie
(2) [+] a great movie good ~~movie~~
(3) [−] poor acting
(4) [+] great acting a good movie

Feature sets for + class documents:

(0)	1	0	0	0	0	0	1	1	0	1
(2)	0	1	0	1	1	0	0	~~2~~ 1	0	0
(4)	0	1	1	1	1	0	0	1	0	0

Feature sets for − class documents:

(1)	1	0	0	0	0	1	0	1	0	1
(3)	0	0	1	0	0	0	0	0	1	0

$P(x_i \mid +)$

I	~~0.0833~~	0.0870
a	~~0.1250~~	0.1304
acting	~~0.0833~~	0.0870
good	~~0.1250~~	0.1304
great	~~0.1250~~	0.1304
hated	~~0.0417~~	0.0435
liked	~~0.0833~~	0.0870
movie	~~0.2083~~	0.1739
poor	~~0.0417~~	0.0435
the	~~0.0833~~	0.0870

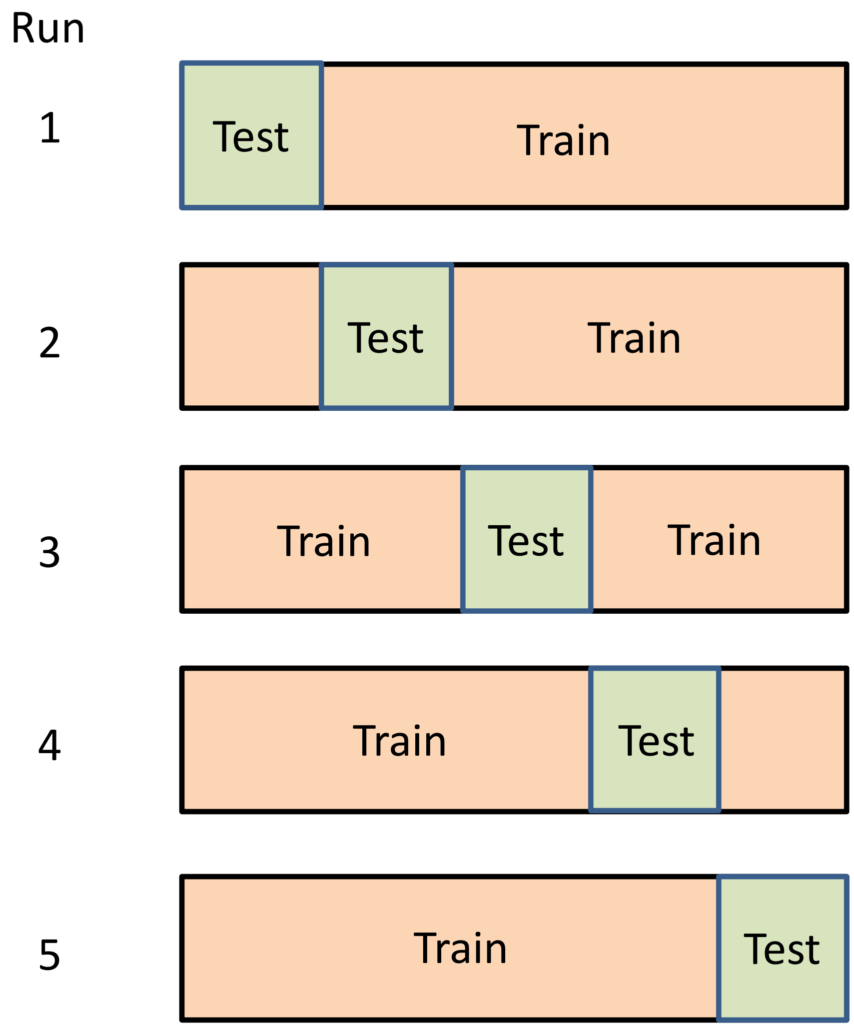
$P(x_i \mid -)$

I	0.1250
a	0.0625
acting	0.1250
good	0.0625
great	0.0625
hated	0.1250
liked	0.0625
movie	0.1250
poor	0.1250
the	0.1250

Example: P(movie | +) = $(\cancel{4}^{\,3}+1) \,/\, (\cancel{14}^{\,13}+10) = \cancel{.2083}^{\,.1739}$

Cross-Validation

- Typically perform 10-fold cross-validation
- Divide data into 10 folds (disjoint subsets)
  - may wish to ensure equal numbers of positive and negative instances
- For each fold
  - train on the other nine
  - test on the fold not trained on
- Report the average performance over the 10 runs



What makes reviews hard to classify?

- Natural language subtlety, metaphors, idioms, and even sarcasm
  - A perfume review from *Perfumes: the Guide:*
    - "If you are reading this because it is your darling fragrance, please wear it at home exclusively, and tape the windows shut."
  - Dorothy Parker (a movie critic) on Katherine Hepburn (an actress)
    - "She runs the gamut of emotions from A to B."

More Reasons

- 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."

Today

- What is Sentiment Analysis?
- A Baseline Algorithm
- Sentiment Lexicons
- Learning Sentiment Lexicons
- Other Sentiment Tasks

Available Sentiment Lexicons (1)

- Harvard General Inquirer (since 1966)
  - labels words simply as positive or negative
    - contains 1916 positive and 2291 negative words
  - based on early work in cognitive psychology and content analysis
  - freely available
  - http://www.wjh.harvard.edu/~inquirer/spreadsheet_guide.htm

1	Entry	Source	Positiv	Negativ	Pstv	Affil	Ngtv	Hostile	Strong	Power	Weak
2	A	H4Lvd									
3	ABANDON	H4Lvd		Negativ			Ngtv				Weak
4	ABANDONMENT	H4		Negativ							Weak
5	ABATE	H4Lvd		Negativ							
6	ABATEMENT	Lvd									
7	ABDICATE	H4		Negativ							Weak
8	ABHOR	H4		Negativ				Hostile			
9	ABIDE	H4	Positiv			Affil					
10	ABILITY	H4Lvd	Positiv						Strong		
11	ABJECT	H4		Negativ							Weak
12	ABLE	H4Lvd	Positiv		Pstv				Strong		
13	ABNORMAL	H4Lvd		Negativ			Ngtv				
14	ABOARD	H4Lvd									
15	ABOLISH	H4Lvd		Negativ			Ngtv	Hostile	Strong	Power	
16	ABOLITION	Lvd									
17	ABOMINABLE	H4		Negativ					Strong		
18	ABORTIVE	Lvd									
19	ABOUND	H4	Positiv								
20	ABOUT#1	H4Lvd									
21	ABOUT#2	H4Lvd									

BigDict1a.tabs

Available Sentiment Lexicons (2)

- Bing Liu's Opinion Lexicon
  - simple lists of 2006 positive and 4783 negative words
  - https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html
  - contains misspelled words that appear frequently in social media
  - contains words that appear in both positive and negative contexts

Excerpts

positive	*negative*
adaptable	abrasive
adaptive	abrupt
adequate	abruptly
adjustable	abscond
admirable	absence
admirably	absent-minded
admiration	absentee
admire	absurd
admirer	absurdity
admiring	absurdly
admiringly	absurdness
adorable	abuse
adore	abused
adored	abuses
adorer	abusive
adoring	abysmal
adoringly	abysmally
adroit	abyss
adroitly	accidental
adulate	accost
adulation	accursed
adulatory	accusation

Available Sentiment Lexicons (3)

- MPQA (Multi-Perspective Question Answering) Subjectivity Lexicon
  - maintained by Theresa Wilson, Janyce Weibe, and Paul Hoffman
  - contains 2718 positive and 4912 negative words from multiple sources
  - words are also labeled for reliability (strongly or weakly subjective)
  - distributed under GNU Public License
  - http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/
  - file downloads with .tff extension – just add .txt to read as text file

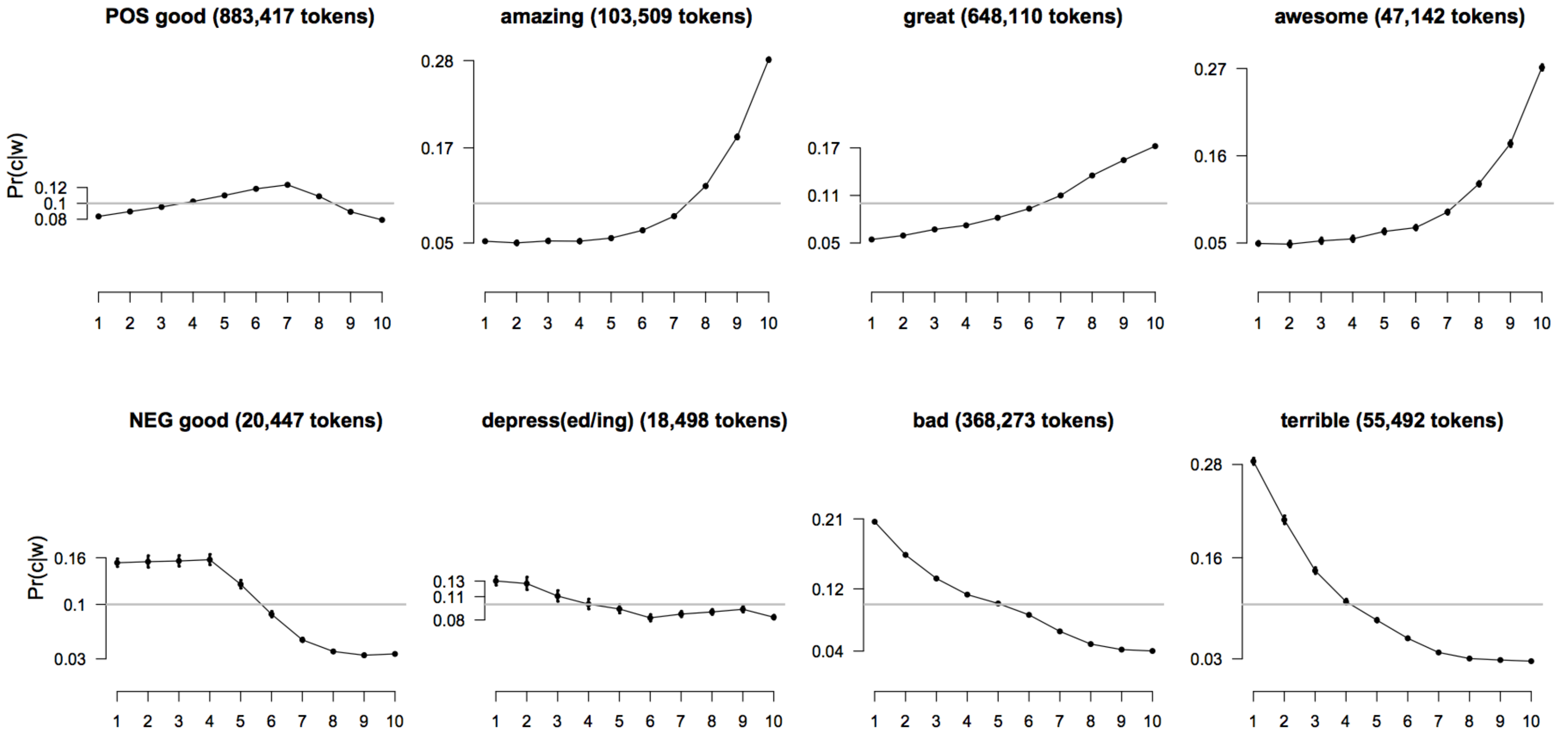
```
type=strongsubj len=1 word1=admonisher pos1=anypos stemmed1=y priorpolarity=negative
type=strongsubj len=1 word1=admonishingly pos1=anypos stemmed1=y priorpolarity=negative
type=strongsubj len=1 word1=admonishment pos1=anypos stemmed1=y priorpolarity=negative
type=strongsubj len=1 word1=admonition pos1=noun stemmed1=n priorpolarity=negative
type=strongsubj len=1 word1=adolescents pos1=adj stemmed1=n priorpolarity=neutral
type=strongsubj len=1 word1=adolescents pos1=noun stemmed1=n priorpolarity=neutral
type=strongsubj len=1 word1=adorable pos1=adj stemmed1=n priorpolarity=positive
type=strongsubj len=1 word1=adore pos1=verb stemmed1=y priorpolarity=positive
type=strongsubj len=1 word1=adored pos1=adj stemmed1=n priorpolarity=positive
type=strongsubj len=1 word1=adorer pos1=anypos stemmed1=y priorpolarity=positive
type=strongsubj len=1 word1=adorer pos1=noun stemmed1=n priorpolarity=positive
type=strongsubj len=1 word1=adoring pos1=adj stemmed1=n priorpolarity=positive
```

How to analyze the polarity of a word

- **Q:** How likely is each word to appear in each sentiment class?

- Example:
  - We can count how many times "bad" appears in 1-star, 2-star, 3-star, etc. reviews
  - But can't use raw counts
  - Instead, we use **likelihood:** $P(w_i|c) = \dfrac{f(wi, c)}{\sum_{w \in c} f(w|c)}$
  - And to compare words, use **scaled likelihood:** $\dfrac{P(w|c)}{P(w)}$

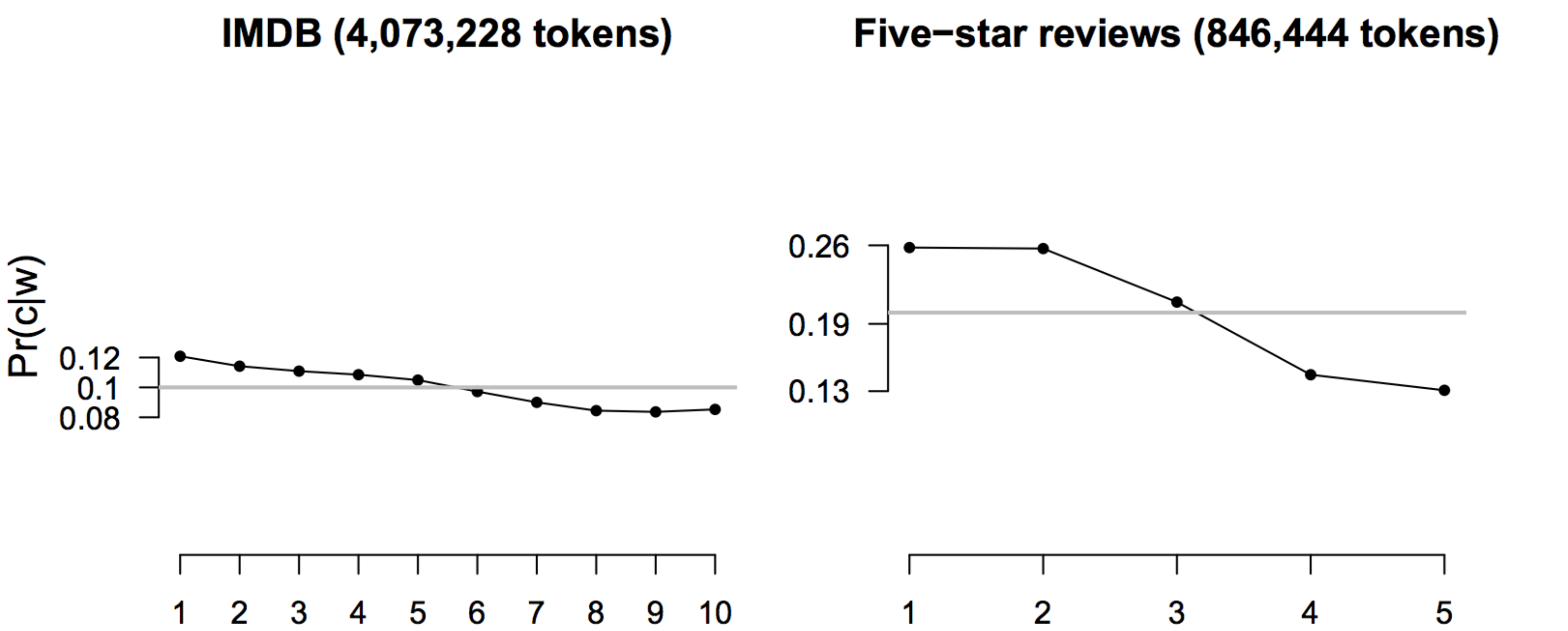
Analyzing word polarity in IMDb



- Source: On the Negativity of Negation, Christopher Potts, SALT 20, 636-659 (2011)

Logical Negation

- Potts (2011) also examined whether negation is associated with negative sentiment
  - Based on counts of negation ("not", "n't", "no", "never") in online reviews
  - Performed regression analysis against the review rating



Today

- What is Sentiment Analysis?
- A Baseline Algorithm
- Sentiment Lexicons
- Learning Sentiment Lexicons
- Other Sentiment Tasks

Semi-supervised learning

- Use **bootstrapping** to build up a lexicon
  - Start with a small amount of information (the "seed" set)
    - a few labeled examples (the "supervised" part)
    - a few hand-built patterns
  - Use the patterns to find related examples
    - use these to augment the original set
    - these additional exemplars were not labeled originally
    - hence the "semi-" in semi-supervised

Identifying word polarity using CCs

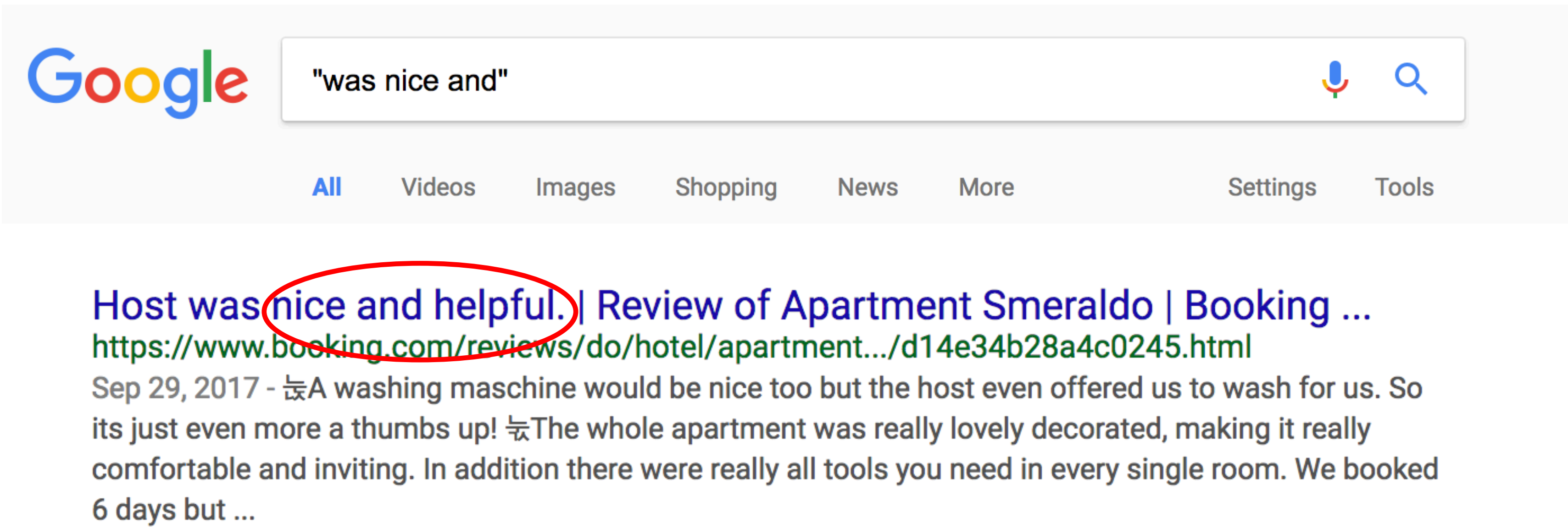
- A "CC" is a standard part-of-speech tag for a "coordinating conjunction"
  - and
  - but
- "Predicting the Semantic Orientation of Adjectives," Hatzivassiloglou and McKeown, ACL 174-181 (1997)
  - intuition
    - detect similar and polar opposite adjectives when conjoined by a CC
    - adjectives conjoined by "and" have same polarity
      - fair **and** legitimate
      - corrupt **and** brutal
    - adjectives conjoined by "but" have opposite polarity
      - fair **but** brutal

Hatzivassiloglou and McKeown – Step 1

- From 21-million word WSJ corpus, extracted 1336 adjectives that occurred > 20 times
- 657 positive
  - adequate central clever famous intelligent remarkable reputed sensitive slender thriving ...
- 679 negative
  - contagious drunken ignorant lanky listless primitive strident troublesome unresolved unsuspecting ...

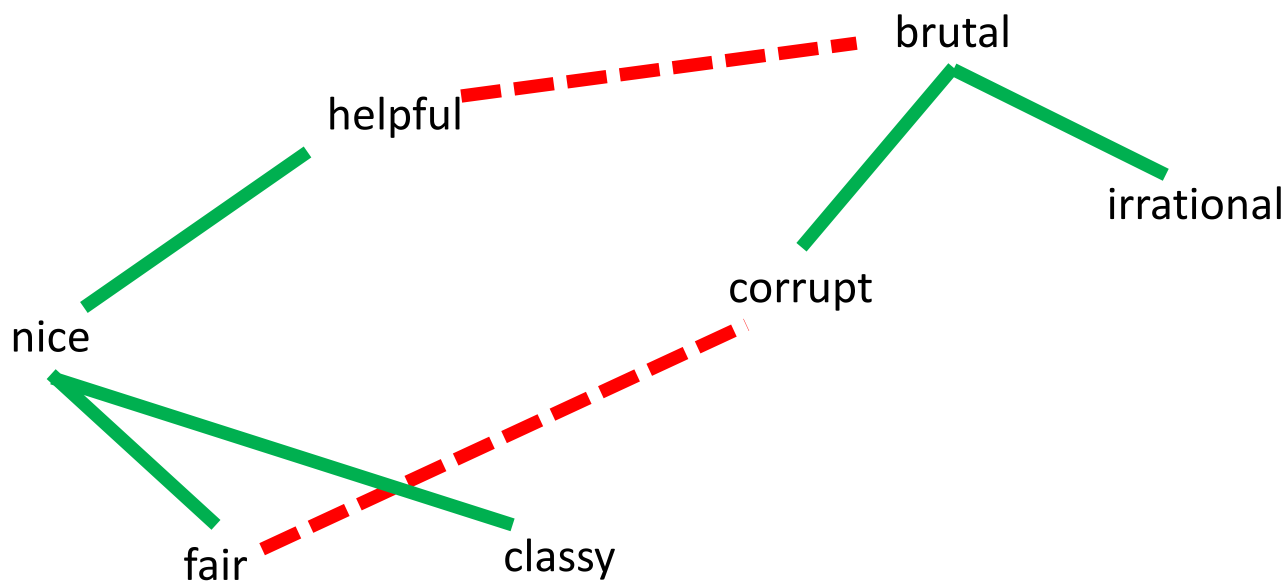
Hatzivassiloglou and McKeown – Step 2

- Expand seed set to conjoined adjectives



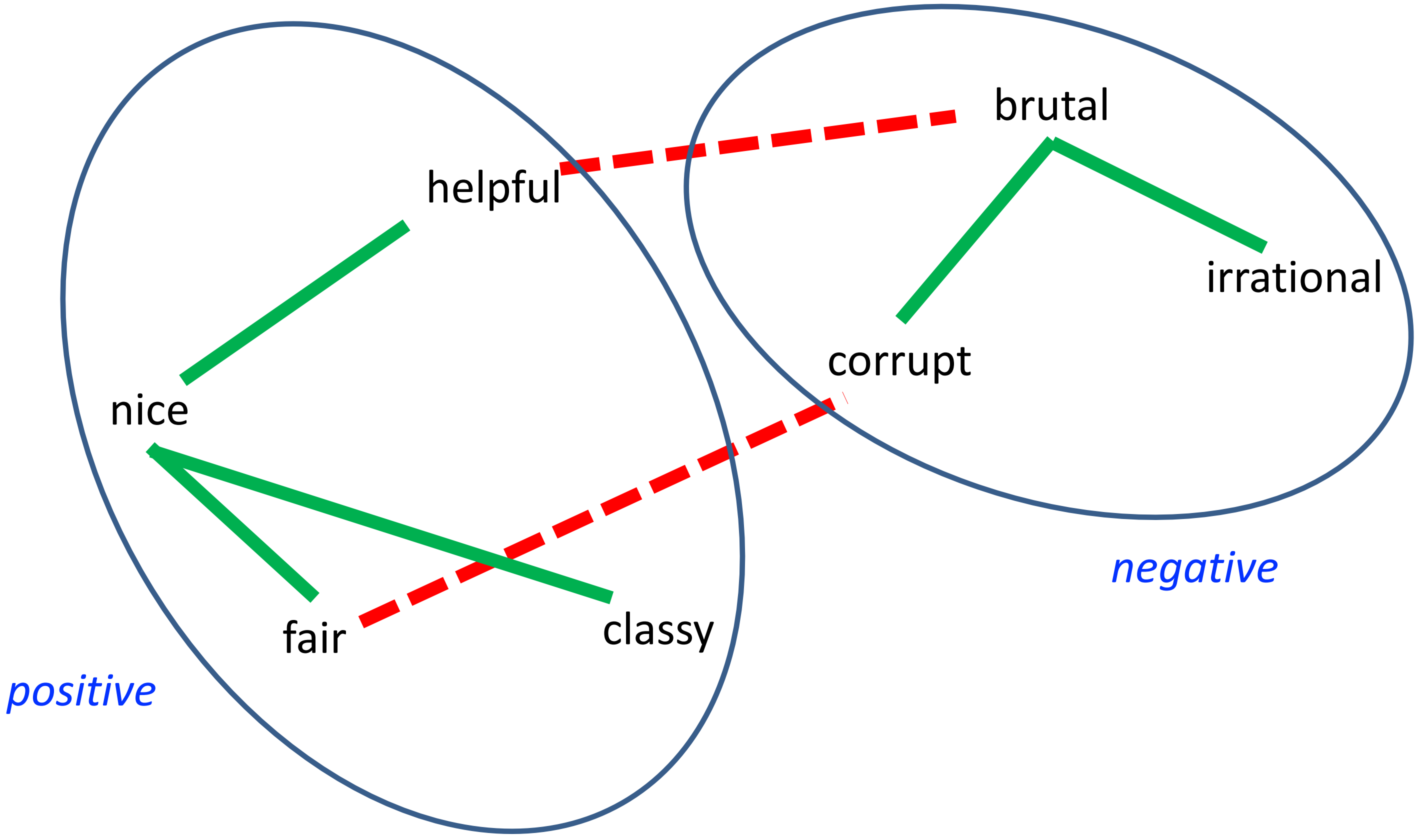
Hatzivassiloglou and McKeown – Step 3

- Supervised classifier assigns "polarity similarity" to each word pair and generates a graph:



Hatzivassiloglou and McKeown – Step 4

- Cluster to partition the graph into positive and negative classes



Turney Algorithm (2002)

- Turney algorithm
  - Extract a phrasal lexicon from reviews
  - Learn polarity of each phrase
  - Rate a review by the average polarity of its phrases

- "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," P. Turney, ACL (2002)

Turney Algorithm – Step 1

- Extract two-word phrases containing adjectives

First Word	Second Word	Third Word (not extracted)	Example
JJ	NN or NNS	anything	online experience
RB, RBR, RBS	JJ	not NN or NNS	very handy
JJ	JJ	not NN or NNS	local online
NN or NNS	JJ	not NN or NNS	programs such
RB, RBR, RBS	VB, VBD, VBN, VBG	anything	inconveniently located

Note: Part-of-speech tags use the Penn Treebank tagset

The polarity of a phrase

- Turney's intuition
  - Positive phrases co-occur more often with "excellent"
  - Negative phrases co-occur more often with "poor"
- Q: But how do we measure these values?

Pointwise Mutual Information

- **Mutual information** between two random variables X and Y:

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

- **Pointwise mutual information** measures how much more, if at all, events X and Y co-occur than if they were independent:

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

- PMI between two words:

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

Estimating PMI

- query a search engine
  - estimate `P(word)` by `hits(word) / N`, where N is size of corpus
  - estimate $\mathtt{P(word_1,\ word_2)}$ by $\mathtt{hits(word_1\ NEAR\ word_2)\ /\ N^2}$

$$PMI(word_1, word_2) = log_2 \frac{hits(word_1\ NEAR\ \ word_2)}{hits(word_1)\ hits(word_2)}$$

  - Note how we don't need to know N, since it cancels out

Co-occurrences with "excellent" and "poor"

- Polarity(phrase) = PMI(phrase, "excellent") – PMI(pharase, "poor")

$$= log_2 \frac{hits(word_1\ NEAR\ \ "excellent")}{hits(word_1)\ hits("excellent")} - log_2 \frac{hits(word_1\ NEAR\ \ "poor")}{hits(word_1)\ hits("poor")}$$

$$= log_2 \frac{hits(word_1\ NEAR\ \ "excellent")}{hits(word_1)\ hits("excellent")} \frac{hits(word_1)\ hits("poor")}{hits(word_1\ NEAR\ \ "poor")}$$

$$= log_2 \left(\frac{hits(word_1\ NEAR\ \ "excellent")}{hits(word_1)\ hits("excellent")} \frac{hits(word_1)\ hits("poor")}{hits(word_1\ NEAR\ \ "poor")} \right)$$

Some Results from Turney (2002)

Table 2. An example of the processing of a review that the author has classified as *recommended*.[6]

Extracted Phrase	Part-of-Speech Tags	Semantic Orientation
online experience	JJ NN	2.253
low fees	JJ NNS	0.333
local branch	JJ NN	0.421
small part	JJ NN	0.053
online service	JJ NN	2.780
printable version	JJ NN	-0.705
direct deposit	JJ NN	1.288
well other	RB JJ	0.237
inconveniently located	RB VBN	-1.541
other bank	JJ NN	-0.850
true service	JJ NN	-0.732
Average Semantic Orientation		0.322

Table 3. An example of the processing of a review that the author has classified as *not recommended*.

Extracted Phrase	Part-of-Speech Tags	Semantic Orientation
little difference	JJ NN	-1.615
clever tricks	JJ NNS	-0.040
programs such	NNS JJ	0.117
possible moment	JJ NN	-0.668
unethical practices	JJ NNS	-8.484
low funds	JJ NNS	-6.843
old man	JJ NN	-2.566
other problems	JJ NNS	-2.748
probably wondering	RB VBG	-1.830
virtual monopoly	JJ NN	-2.050
other bank	JJ NN	-0.850
extra day	JJ NN	-0.286
direct deposits	JJ NNS	5.771
online web	JJ NN	1.936
cool thing	JJ NN	0.395
very handy	RB JJ	1.349
lesser evil	RBR JJ	-2.288
Average Semantic Orientation		-1.218

Today

- What is Sentiment Analysis?
- A Baseline Algorithm
- Sentiment Lexicons
- Learning Sentiment Lexicons
- Other Sentiment Tasks

Sentiment Attributes

- Consider the sentence:

"The food was great but the service was terrible."

➔ It is important to know to what "great" and "terrible" refer

Using Frequent Phrases

- Find all frequent phrases across the reviews
  - looking for unknown items
    - e.g., aspects of a movie, hotel, particular dishes at a restaurant
  - issue of thresholding
  - example phrase: fish tacos

- use word-based rules to extract sentiment words
  - e.g.,look for <sentiment word> immediately before <freq_phrase>
  - e.g., "great fish tacos" ⟶ "fish tacos" a likely ***aspect*** of sentiment

- Similarly, one can find aspects for
  - restaurant – food, service, wine selection, lamb, appetizer
  - hotel – parking, price, pool, buffet

Remote Aspects

- Sometimes, the aspects are not in the sentences that express sentiment

- For certain problems, the aspects are well-understood
  - i.e., known ahead of time
  - e.g., for hotels and restaurants

- Approach
  - Hand-label a small corpus of review with the known aspects
    - e.g., for restaurant: food, decor, service, value, NONE

  - Train a classifier to assign aspect to a given sentence
    - NONE needed, since not all sentences will express sentiment on predefined aspect list
    - e.g., for hotel: "It was a long taxi ride to the hotel from the airport."

Other Research Areas in Affective Computing

- Sentiment is only one aspect of affective computing. Other research areas (identified by Scherer's topology) include
  - Emotion
    - detecting annoyed callers in a dialogue system
    - detecting confused or frustrated students, versus confident ones
  - Mood
    - detecting traumatized or depressed writers
  - Interpersonal stances
    - detecting friendliness or hostility in conversations
  - Personality traits
    - detecting extroverts and introverts