Sentiment Analysis - Introduction

Pawan Goyal

CSE, IIT Kharagpur

Week 12, Lecture 1

Example: 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.

Where is Sentiment Analysis Used?

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 marked trends from sentiment

Where is Sentiment Analysis Used?

- Frustration of callers to a help line
- Stress in drivers or pilots
- Depression and other medical conditions from social media
- Confusion in students talking to e-tutors

Affective States Typology

Emotion: angry, sad, joyful, fearful, ashamed, proud, elated

Affective States Typology

Emotion: angry, sad, joyful, fearful, ashamed, proud, elated

Mood: cheerful, gloomy, irritable, listless, depressed, buoyant

Affective States Typology

Emotion: angry, sad, joyful, fearful, ashamed, proud, elated

Mood: cheerful, gloomy, irritable, listless, depressed, buoyant

Interpersonal stances: friendly, flirtatious, distant, cold, warm, supportive, contemptuous

Affective States Typology

Emotion: angry, sad, joyful, fearful, ashamed, proud, elated

Mood: cheerful, gloomy, irritable, listless, depressed, buoyant

Interpersonal stances: friendly, flirtatious, distant, cold, warm, supportive, contemptuous

Attitudes: liking, loving, hating, valuing, desiring

Affective States Typology

Emotion: angry, sad, joyful, fearful, ashamed, proud, elated

Mood: cheerful, gloomy, irritable, listless, depressed, buoyant

Interpersonal stances: friendly, flirtatious, distant, cold, warm, supportive, contemptuous

Attitudes: liking, loving, hating, valuing, desiring

Personality Traits: nervous, anxious, reckless, morose, hostile, jealous

Affective States Typology

Emotion: angry, sad, joyful, fearful, ashamed, proud, elated

Mood: cheerful, gloomy, irritable, listless, depressed, buoyant

Interpersonal stances: friendly, flirtatious, distant, cold, warm, supportive, contemptuous

Attitudes: *liking, loving, hating, valuing, desiring*

Personality Traits: nervous, anxious, reckless, morose, hostile, jealous

Sentiment Analysis

Sentiment Analysis is the detection of attitudes

enduring, affectively colored beliefs, dispositions towards objects or persons

Sentiment Analysis

Sentiment Analysis is the detection of attitudes

enduring, affectively colored beliefs, dispositions towards objects or persons

The complete task

- Holder (source) of attitude
- Target (aspect) of attitude
- Type of attitude
  - From a set of types: *like, love, hate, value, desire*
  - Or simple weighted polarity: *positive, negative, neutral, together with strength*
- Text containing the attitude

Simplest Task

Is the attitude of this text positive or negative?

Simplest Task

Is the attitude of this text positive or negative?

More complex

Rank the attitudes of this text from 1 to 5

Simplest Task

Is the attitude of this text positive or negative?

More complex

Rank the attitudes of this text from 1 to 5

Advanced

Detect the target, source, or complex attitude types

Sentiment Analysis in Movie Reviews

Polarity detection

Is an IMDB movie review positive or negative?

✓

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

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

Tokenization Issues

- Capitalization - preserve for word in all caps
- Word lengthening
- Handling emoticons

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

- Handling negation
  - I **didn't** like this movie
  - I really like this movie

Add *NOT_* to every word between negation and following punctuation

  - *didn't like this movie, but I ...*

Tokenization Issues

- Capitalization - preserve for word in all caps
- Word lengthening
- Handling emoticons

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

- Handling negation
  - I **didn't** like this movie
  - I really like this movie

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..

Naïve Bayes: Reminder

$$c_{NB} = \operatorname*{arg\,max}_{c_j \in C} P(c_j) \prod_{x_i} P(x_i|c_j)$$

$$\hat{P}(c_j) = \frac{doc-count(C = c_j)}{N_{doc}}$$

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

Boolean Multinomial Naïve Bayes

- First remove all duplicate words from a test document d
- Then compute NB using the same equation

$$c_{NB} = \arg\max_{c_j \in C} P(c_j) \prod_{x_i} P(x_i|c_j)$$

A piece of cake?

Is a given review on a known topic positive or negative?

> "It may be a bit early to make such judgments, but Battlefield Earth may well turn out to be the worst movie of this century." (Elvis Mitchell, May 12, 2000)

A piece of cake?

Is a given review on a known topic positive or negative?

"It may be a bit early to make such judgments, but Battlefield Earth may well turn out to be the worst movie of this century." (Elvis Mitchell, May 12, 2000)

don't we just need to look for "worst", "best", "love", hate", etc.?

In a small scale experiment (Pang et al., 2002)

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

Why can't we just look for words like "great" and "terrible"?

- This laptop is *a great deal*.

Why can't we just look for words like "great" and "terrible"?

- This laptop is *a great deal*.
- *A great deal* of media attention surrounded the release of the new laptop.

Why can't we just look for words like "great" and "terrible"?

- 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.

Why can't we just look for words like "great" and "terrible"?

- 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.
- 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.

Sentiment Analysis - Affective Lexicons

Pawan Goyal

CSE, IIT Kharagpur

Week 12, Lecture 2

Sentiment Lexicons

- The general inquirer
- MPQA Subjectivity Cues Lexicon
- SentiWordnet
- LIWC (Linguistic Inquiry and Word Count)

The General Inquirer

Categories

- Positive (1915 words) and Negative (2291) words
- Strong vs weak, active vs passive, overstated vs understated
- pleasure, pain, virtue, vice, motivation, cognitive orientation etc.

Entry	Source	Positiv	Negativ	Pstv	Affil	Ngtv	Hostile	Strong	Power	Weak	Submit	Active	Pa
A	H4Lvd												
ABANDON	H4Lvd		Negativ			Ngtv				Weak			
ABANDONMENT	H4		Negativ							Weak			
ABATE	H4Lvd		Negativ										Pa
ABATEMENT	Lvd												
ABDICATE	H4		Negativ							Weak	Submit		Pa
ABHOR	H4		Negativ				Hostile						Pa
ABIDE	H4	Positiv			Affil							Active	
ABILITY	H4Lvd	Positiv						Strong					
ABJECT	H4		Negativ							Weak	Submit		Pa
ABLE	H4Lvd	Positiv		Pstv				Strong					
ABNORMAL	H4Lvd		Negativ			Ngtv							
ABOARD	H4Lvd												
ABOLISH	H4Lvd		Negativ			Ngtv	Hostile	Strong	Power			Active	
ABOLITION	Lvd												
ABOMINABLE	H4		Negativ					Strong					
ABORTIVE	Lvd												
ABOUND	H4	Positiv											Pa
ABOUT#1	H4Lvd												
ABOUT#2	H4Lvd												
ABOUT#3	H4Lvd												
ABOUT#4	H4Lvd												
ABOUT#5	H4Lvd												
ABOUT#6	H4Lvd												
ABOUT#7	H4Lvd												
ABOVE#1	H4Lvd												
ABOVE#2	H4Lvd												
ABOVE#3	H4Lvd												
ABOVE#4	H4Lvd												

SentiWordNet

- Home page: `http://sentiwordnet.isti.cnr.it/`
- All Wordnet synsets automatically annotated for degrees of positivity, negativity, and neutrality/objectiveness

SentiWordNet

- Home page: `http://sentiwordnet.isti.cnr.it/`
- All Wordnet synsets automatically annotated for degrees of positivity, negativity, and neutrality/objectiveness

Example

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

Other Lexicons

MPQA Subjectivity Cues Lexicon

- 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)

Other Lexicons

MPQA Subjectivity Cues Lexicon

- 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)

Bing Liu Opinion Lexicon

- Bing Liu's Page on Opinion Mining
- http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar
- 6786 word: 2006 positive, 4780 negative

LIWC (Linguistic Inquiry and Word Count)

- Home page: `http://www.liwc.net/`
- 2300 words, > 70 classes

LIWC (Linguistic Inquiry and Word Count)

- Home page: `http://www.liwc.net/`
- 2300 words, > 70 classes

Affective Processes

- Negative emotion (bad, weird, hate, problem, tough)
- Positive emotion (love, nice, sweet)

LIWC (Linguistic Inquiry and Word Count)

- 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)

LIWC (Linguistic Inquiry and Word Count)

- Home page: `http://www.liwc.net/`
- 2300 words, > 70 classes

Affective Processes

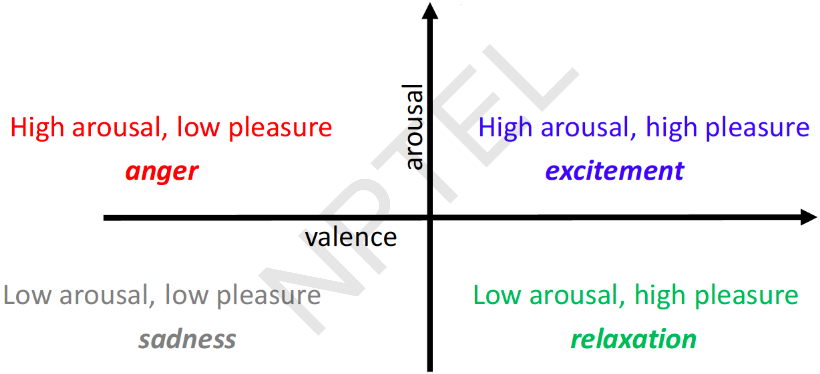
- Negative emotion (bad, weird, hate, problem, tough)
- Positive emotion (love, nice, sweet)

Cognitive Processes

Tentative (maybe, perhaps, guess), Inhibition (block, constraint)

- Comes with a small fee

Valence / Arousal Dimenstions



Lexicon of valence, arousal, and dominance

- Warriner, Amy Beth, Victor Kuperman, and Marc Brysbaert. "Norms of valence, arousal, and dominance for 13,915 English lemmas." Behavior research methods 45.4 (2013): 1191-1207.
- **Supplementary data:** This word is licenced under a Creative Commons.

Lexicon of valence, arousal, and dominance

- Warriner, Amy Beth, Victor Kuperman, and Marc Brysbaert. "Norms of valence, arousal, and dominance for 13,915 English lemmas." Behavior research methods 45.4 (2013): 1191-1207.
- **Supplementary data:** This word is licenced under a Creative Commons.

Ratings for 14,000 words for emotional dimensions:

- **valence** (the pleasantness of the stimulus)
- **arousal** (the intensity of emotion provoked by the stimulus)
- **dominance** (the degree of control exerted by the stimulus)

Lexicon of valence, arousal, and dominance

valence (the pleasantness of the stimulus

- 9: happy, pleased, satisfied, contented, hopeful
- 1: unhappy, annoyed, unsatisfied, melancholic, despaired, or bored

Lexicon of valence, arousal, and dominance

valence (the pleasantness of the stimulus

- 9: happy, pleased, satisfied, contented, hopeful
- 1: unhappy, annoyed, unsatisfied, melancholic, despaired, or bored

arousal (the intensity of emotion provoked by the stimulus

- 9: stimulated, excited, frenzied, jittery, wide-awake, or aroused
- 1: relaxed, calm, sluggish, dull, sleepy, or unaroused

Lexicon of valence, arousal, and dominance

valence (the pleasantness of the stimulus

- 9: happy, pleased, satisfied, contented, hopeful
- 1: unhappy, annoyed, unsatisfied, melancholic, despaired, or bored

arousal (the intensity of emotion provoked by the stimulus

- 9: stimulated, excited, frenzied, jittery, wide-awake, or aroused
- 1: relaxed, calm, sluggish, dull, sleepy, or unaroused

dominance (the degree of control exerted by the stimulus)

- 9: in control, influential, important, dominant, autonomous, or controlling
- 1: controlled, influenced, cared-for, awed, submissive, or guided

Lexicon of valence, arousal, and dominance: Examples

Valence		Arousal		Dominance	
vacation	8.53	rampage	7.56	self	7.74
happy	8.47	tornado	7.45	incredible	7.74
whistle	5.7	zucchini	4.18	skillet	5.33
conscious	5.53	dressy	4.15	concur	5.29
torture	1.4	dull	1.67	earthquake	2.14

Learning Affective Lexicons

Pawan Goyal

CSE, IIT Kharagpur

Week 12, Lecture 3

Learning Sentiment Lexicons

Basic Intuition

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

Learning Sentiment Lexicons

*Step 1: Label **seed set** of adjectives*

Learning Sentiment Lexicons

Step 1: Label ***seed set*** *of adjectives*

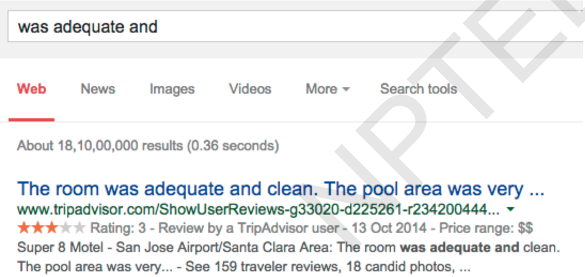
- **Positive cases:** adequate, central, clever, famous, intelligent, remarkable, reputed, sensitive, slender, thriving ...
- **Negative cases:** contagious, drunken, ignorant, lanky, listless, primitive, strident, troublesome, unresolved, unsuspecting ...

Learning Sentiment Lexicons

Step 2: Expand seed set to conjoined adjectives

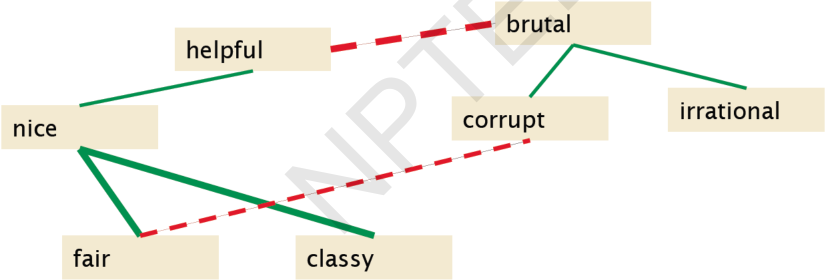
Learning Sentiment Lexicons

Step 2: Expand seed set to conjoined adjectives

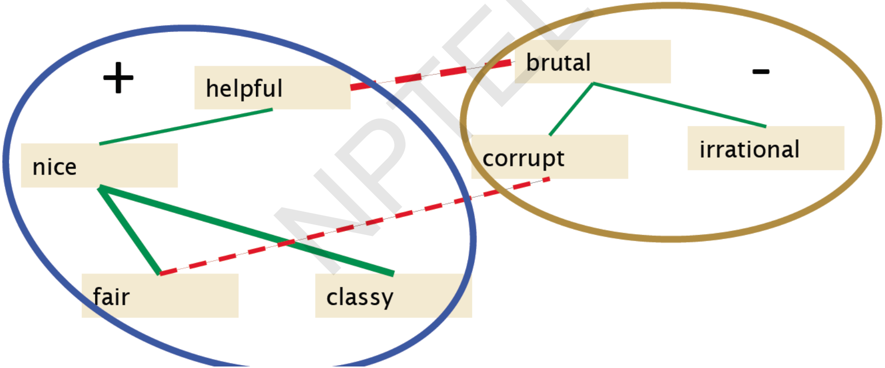


Step 3: Construct a graph

Polarity similarity is assigned to each word pair:



Clustering for partitioning the graph into two



Output Polarity Lexicon

Positive

bold decisive disturbing generous good honest important large mature patient peaceful positive proud sound stimulating straightforward strange talented vigorous witty ...

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

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

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

Measuring the polarity of the phrases

Measuring the polarity of the phrases

- Positive phrases co-occur more with “excellent”
- Negative phrases co-occur more with “poor”
- How to measure the co-occurrence?

Measuring the polarity of the phrases

- Positive phrases co-occur more with "excellent"
- Negative phrases co-occur more with "poor"
- How to measure the co-occurrence?

Pointwise Mutual Information

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

How to estimate PMI?

Query search engine (Altavista)

How to estimate PMI?

Query search engine (Altavista)

- $P(word)$ estimated by $hits(word)$/N
- $P(word_1, word_2)$ estimated by $hits(word_1$ NEAR $word_2)$/N

How to estimate PMI?

Query search engine (Altavista)

- $P(word)$ estimated by $hits(word)$/N
- $P(word_1, word_2)$ estimated by $hits(word_1$ NEAR $word_2)$/N

Polarity(phrase) = PMI(phrase, excellent) - PMI(phrase, poor)

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

Example: 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

Example: 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

Using WordNet to learn polarity

- WordNet: online thesaurus
- 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

Computing with Affective Lexicons

Pawan Goyal

CSE, IIT Kharagpur

Week 12, Lecture 4

Learn word sentiment supervised by online review scores

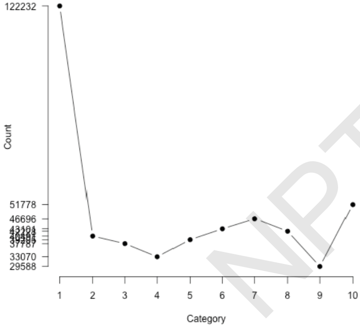
- Review datasets: IMDB, Goodreads, Amazon, Trip Advisor
- Each review has a score (1-5, 1-10 etc)
- Just count how many times each word occurs with each score (and normalize).

Analyzing polarity of each word in IMDB

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

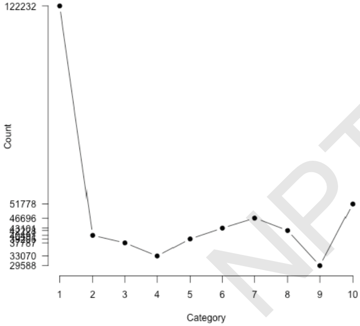
Analyzing polarity of each word in IMDB

- How likely is each word to appear in each sentiment class?
- Let’s take count(“bad”) in 1-star, 2-star, 3-star etc.



Analyzing polarity of each word in IMDB

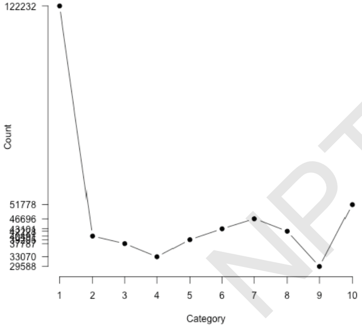
- How likely is each word to appear in each sentiment class?
- Let's take count("bad") in 1-star, 2-star, 3-star etc.



- We should use likelihood instead of counts: $P(w|c) = \frac{f(w,c)}{\sum_{w \in c} f(w,c)}$

Analyzing polarity of each word in IMDB

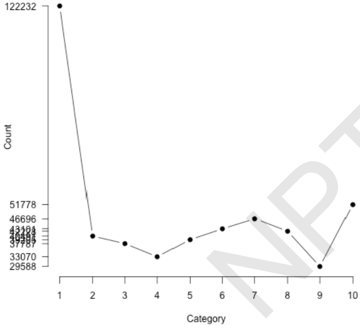
- How likely is each word to appear in each sentiment class?
- Let's take count("bad") in 1-star, 2-star, 3-star etc.



- We should use likelihood instead of counts: $P(w|c) = \frac{f(w,c)}{\sum_{w \in c} f(w,c)}$
- Make them comparable between words

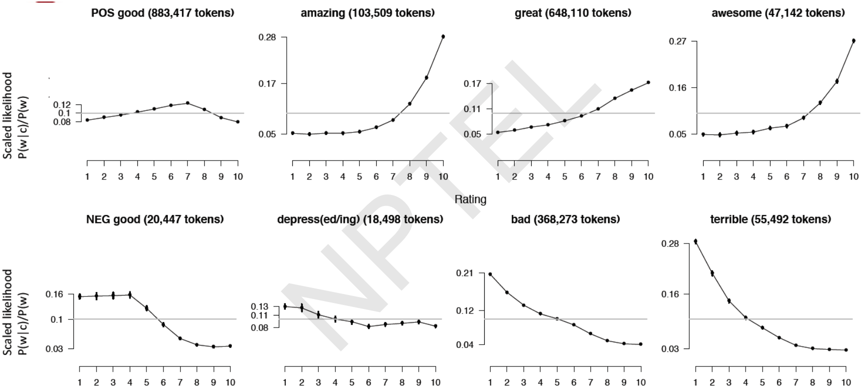
Analyzing polarity of each word in IMDB

- How likely is each word to appear in each sentiment class?
- Let's take count("bad") in 1-star, 2-star, 3-star etc.



- We should use likelihood instead of counts: $P(w|c) = \frac{f(w,c)}{\sum_{w \in c} f(w,c)}$
- Make them comparable between words Scaled likelihood: $\frac{P(w|c)}{P(w)}$

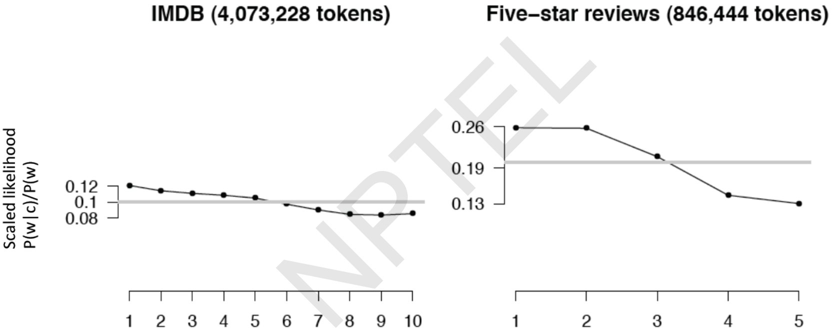
Analyzing polarity in IMDB



Logical Negation

- 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

More negation in negative sentiment



Using Linguistic Intuitions

Using a sentiment lexicon also works.

Using Linguistic Intuitions

Using a sentiment lexicon also works.
Some linguistic intuitions on top of that tends to give better results.

Handling negation in simple addition of scores

Example words

- Excellent +5
- good +3
- terrible -5
- bad -3

Handling negation in simple addition of scores

Example words

- Excellent +5
- good +3
- terrible -5
- bad -3

Reversing the polarity

- Not Excellent -5
- Not good -3
- Not terrible +5
- Not bad +3

Handling negation in simple addition of scores

Example words

- Excellent +5
- good +3
- terrible -5
- bad -3

Handling negation in simple addition of scores

Example words

- Excellent +5
- good +3
- terrible -5
- bad -3

Instead, a polarity shift works better

- Not Excellent (5-4) +1
- Not good (3-4) -1
- Not terrible (-5+4) -1
- Not bad (-3+4) 1

Handling Intensifiers

Intensifiers can be classified into two major categories,

- Amplifiers (e.g., very) increase the semantic intensity
- Downtoners (e.g., slightly) decrease it

Handling Intensifiers

Intensifiers can be classified into two major categories,

- Amplifiers (e.g., very) increase the semantic intensity
- Downtoners (e.g., slightly) decrease it

Rough values for some intensifiers

Intensifier	Modifier (%)
slightly	−50
somewhat	−30
pretty	−10
really	+15
very	+25
extraordinarily	+50
(the) most	+100

Handling Intensifiers

Intensifiers can be classified into two major categories,

- Amplifiers (e.g., very) increase the semantic intensity
- Downtoners (e.g., slightly) decrease it

Rough values for some intensifiers

Intensifier	Modifier (%)
slightly	−50
somewhat	−30
pretty	−10
really	+15
very	+25
extraordinarily	+50
(the) most	+100

Somewhat sleazy

sleazy: -3, *somewhat sleazy*: $-3 \times (100\% - 30\%) = -2.1$

Irrealis moods: where the words may not be reliable

- I thought this movie would be as good as the Grinch, but unfortunately, it wasn't.
- This should have been a great movie.

Irrealis moods: where the words may not be reliable

- I thought this movie would be as good as the Grinch, but unfortunately, it wasn't.
- This should have been a great movie.

What are the indicators?

- conditional markers (if)
- negative polarity items like 'any' and 'anything'
- certain (mostly intensional) verbs (*expect, doubt*),
- questions
- words enclosed in quotes (which may be factual, but not necessarily reflective of the author's opinion)

Text scoring

*Shows how a variety of **sentiment lexicons** score novel texts. Such values could be used in many ways (as raw values, to derive percentages or ratios, as **classifier features**, ...).*

Enter your own text (max 140 characters), or

It sounds like a great plot but can't hold up.

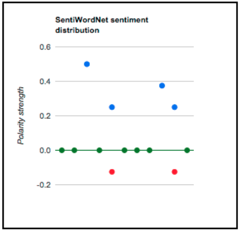
☐ analyze a random tweet instead.

Submit

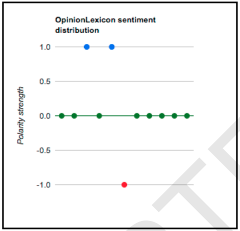
Scores

```
it sounds like a great plot but can't hold_NEG up_NEG .
```

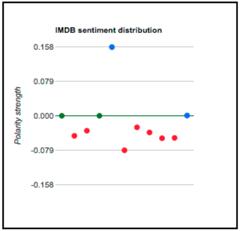
SentiWordNet



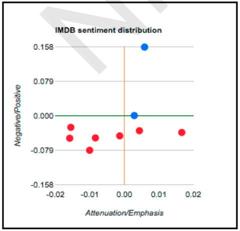
Opinion Lexicon



IMDB



IMDB 2d



Aspect-based Sentiment Analysis

Pawan Goyal

CSE, IIT Kharagpur

Week 12, Lecture 5

Finding aspects or attributes

The food was great but the service was awful.

Finding aspects or attributes

The food was great but the service was awful.

Aspects Involved

Food, service

Finding aspect/attribute/target of sentiment

Finding aspect/attribute/target of sentiment

Frequent phrases + rules

Finding aspect/attribute/target of sentiment

Frequent phrases + rules

- Find all highly frequent phrases across reviews ("fish tacos")

Finding aspect/attribute/target of sentiment

Frequent phrases + rules

- Find all highly frequent phrases across reviews (“fish tacos”)
- Filter by rules like “occurs right after sentiment word”

Finding aspect/attribute/target of sentiment

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

Finding aspect/attribute/target of sentiment

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

Extraction of aspect-opinion pairs

RuleID	Observations	Examples
R1	$JJ \leftarrow amod \leftarrow NP$	The camera has a good screen. (good ← amod ← screen)
R2	$NP \rightarrow nsubj \rightarrow JJ$	The flash is brilliant. (flash ← nsubj ← brilliant)
R3	$VB \rightarrow dobj \rightarrow NP$	I love the image quality. (love → dobj → image quality)
R4	$NP \rightarrow nsubj \rightarrow JJ$, JJ ∈ implicit aspect lexicon	The camera is expensive. (camera ← nsubj ← expensive)

Finding aspect/attribute/target of sentiment

If the aspects are well understood, use supervised classification.

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

Do opinion phrases always have the same sentiment?

'Large' – positive or negative

Large screen vs. Large battery

Do opinion phrases always have the same sentiment?

'Large' – positive or negative

Large screen vs. Large battery

'Long' – positive or negative

Long battery life vs. Long loading time

Explicit vs. Implicit Aspect Expressions

- *The picture quality of this camera is great'* – 'picture quality' is an explicit aspect.
- *This camera is expensive* – 'expensive' is an implicit aspect expression desribing 'price'.
- Implicit aspect expressions can be very complex as well, e.g., *This camera will not fit in a pocket* – "fit in a pocket' indicates the aspect 'size'.

Interesting Applications

Aspect-based Opinion Summarization

"“I bought an iPhone a few days ago. It is such a nice phone. The touch screen is really cool. The voice quality is clear too. It is much better than my old Blackberry, which was a terrible phone and so difficult to type with its tiny keys. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, …”

Feature Based Summary of iPhone:

Feature1: **Touch screen**

Positive: 212

- *The touch screen was really cool.*
- *The touch screen was so easy to use and can do amazing things.*

…

Negative: 6

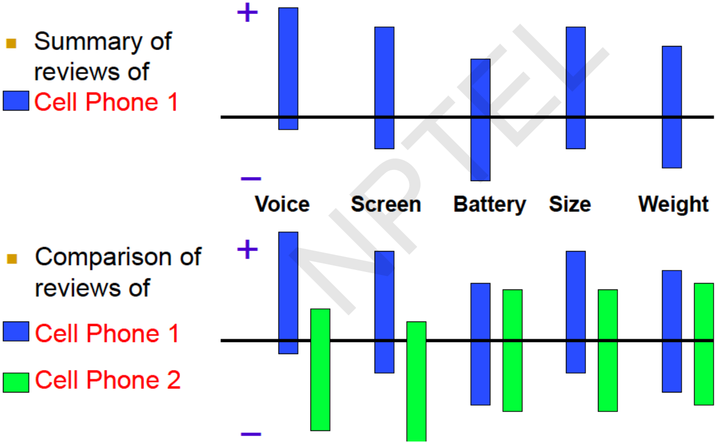
- The screen is easily scratched.
- I have a lot of difficulty in removing finger marks from the touch screen.

…

Feature2: **voice quality**

Interesting Applications

Aspect-based Product Comparison



Conclusion

Many more NLP Applications

- Machine Translation
- Question Answering
- Cross-lingual Applications
- Text Processing for social media, informal text

Conclusion

Many more NLP Applications

- Machine Translation
- Question Answering
- Cross-lingual Applications
- Text Processing for social media, informal text

Deep Learning Techniques are being applied for most of the tasks.