CS 6120/CS4120: Natural Language Processing

Instructor: Prof. Lu Wang
College of Computer and Information Science
Northeastern University
Webpage: www.ccs.neu.edu/home/luwang
Sentiment Analysis
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.
iPhone 8 vs iPhone 6s Plus, iPhone SE, & iPhone X. – December 24, 2017

– Review provided by Best Buy
December 24, 2017

I have recent experience using both the iPhone SE and iPhone 6s Plus. The Plus model was too big since I use a case with a belt clip to carry the phone, and the SE's screen was a bit too small. I am going to compare my review of the iPhone 8 (purchased unlocked at full price and used with Verizon prepaid) mostly to the iPhone 6s Plus, but one has to understand that the SE especially at prepaid price is an excellent, outstanding phone too with almost all of the same features as the Plus!

Pros: The iPhone 8 is an upgrade in a few ways. Apple includes a compare feature on its website so I won’t go into all of the details, but I will try to address the ones that are upgrades to the iPhone 6s Plus. True Tone display does make the screen easier to read because the lighting isn’t ... more »
Twitter sentiment versus Gallup Poll of Consumer Confidence

Twitter sentiment:

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
Sentiment Analysis

• Sentiment analysis is the detection of **attitudes**
  “enduring, affectively colored beliefs, dispositions towards objects or persons”

  *Emily told Charlie that the new movie is disappointing.*

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

• 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 Classification in Movie Reviews


• Polarity detection:
  • Is an IMDB movie review positive or negative?

• Data: *Polarity Data 2.0:*
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 de palma 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](#)
Pre-processing Social Media Text

• Social Media Text is noisy
  • Informal e.g., slangs
  • Misspellings e.g., covfefe
  • Elongated words e.g., can’t waittt
  • Hashtags
  • Emoticons
  •Urls
  • Random capitalization e.g., NOT COOL!

[Borrowed from Kishaloy Halder’s Slides]
Pre-processing: Hashtags

• Hashtagged words are good labels of sentiments and emotions
  • Can’t wait to have my own Google glasses #awesome
  • Some jerk just stole my photo on #tumblr. #grr #anger

• Hashtag Sentiment Lexicon
  • created from a large collection of hashtagged tweets

• New hashtags are being generated every minute

• Breaking long hashtags into smaller instances
  • #killthebill -> kill the bill
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 often turn out to work better
Add NOT_ to every word between negation and following punctuation:

didn’t NOT_like NOT_this NOT_movie, but I

didn’t NOT_like NOT_this NOT_movie but I
Reminder: Naïve Bayes

\[ c_{NB} = \arg\max_{c_j \in C} P(c_j) \prod_{i \in \text{positions}} P(w_i \mid c_j) \]

\[ \hat{P}(w \mid c) = \frac{\text{count}(w, c) + 1}{\text{count}(c) + |V|} \]
Binarized (Boolean feature)

• Intuition:
  • For sentiment (and probably for other text classification domains)
  • Word occurrence may matter more than word frequency
    • The occurrence of the word *fantastic* tells us a lot
    • The fact that it occurs 5 times may not tell us much more.
  • Boolean Multinomial Naïve Bayes
    • Clips all the word counts in each document at 1
## Normal vs. Boolean Multinomial NB

<table>
<thead>
<tr>
<th>Normal</th>
<th>Doc</th>
<th>Words</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>1</td>
<td>Chinese Beijing Chinese</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Chinese Chinese Shanghai</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Chinese Macao</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Tokyo Japan Chinese</td>
<td>j</td>
</tr>
<tr>
<td>Test</td>
<td>5</td>
<td>Chinese Chinese Chinese Tokyo Japan</td>
<td>?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Boolean</th>
<th>Doc</th>
<th>Words</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>1</td>
<td>Chinese Beijing</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Chinese Shanghai</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Chinese Macao</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Tokyo Japan Chinese</td>
<td>j</td>
</tr>
<tr>
<td>Test</td>
<td>5</td>
<td>Chinese Tokyo Japan</td>
<td>?</td>
</tr>
</tbody>
</table>
Binarized (Boolean feature) Multinomial Naïve Bayes

- Binary seems to work better than full word counts
- Other possibility: $\log(\text{freq}(w))$

---


Problems: What makes reviews hard to classify?

• Subtlety:
  • Perfume review in *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 on Katherine Hepburn
    • “She runs the gamut of emotions from A to B”
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.
Sentiment Lexicons
The General Inquirer


- 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)


- 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*)

- Not free though!
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)

• GNU GPL
Bing Liu Opinion Lexicon


- 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/](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

<table>
<thead>
<tr>
<th></th>
<th>Opinion Lexicon</th>
<th>General Inquirer</th>
<th>SentiWordNet</th>
<th>LIWC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MPQA</strong></td>
<td>33/5402 (0.6%)</td>
<td>49/2867 (2%)</td>
<td>1127/4214 (27%)</td>
<td>12/363 (3%)</td>
</tr>
<tr>
<td><strong>Opinion Lexicon</strong></td>
<td>32/2411 (1%)</td>
<td>1004/3994 (25%)</td>
<td></td>
<td>9/403 (2%)</td>
</tr>
<tr>
<td><strong>General Inquirer</strong></td>
<td></td>
<td>520/2306 (23%)</td>
<td>1/204 (0.5%)</td>
<td></td>
</tr>
<tr>
<td><strong>SentiWordNet</strong></td>
<td></td>
<td></td>
<td></td>
<td>174/694 (25%)</td>
</tr>
<tr>
<td><strong>LIWC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Analyzing the polarity of each word in IMDB


• 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:**
    \[
    P(w | c) = \frac{f(w, c)}{\sum_{w \in c} f(w, c)}
    \]
• Make them comparable between words
  • Scaled likelihood:
    \[
    \frac{P(w | c)}{P(w)}
    \]
Analyzing the polarity of each word in IMDB

Other sentiment feature: 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
Potts 2011 Results: More negation in negative sentiment

IMDB (4,073,228 tokens)  Five-star reviews (846,444 tokens)

Scaled likelihood $P(w|c)/P(w)$

1 2 3 4 5 6 7 8 9 10

1 2 3 4 5

0.26
0.19
0.13
0.12
0.1
0.08
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


• 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 have the same polarity
  • 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 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
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

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

Extract two-word phrases with adjectives

<table>
<thead>
<tr>
<th>First Word</th>
<th>Second Word</th>
<th>Third Word (not extracted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JJ</td>
<td>NN or NNS</td>
<td>anything</td>
</tr>
<tr>
<td>RB, RBR, RBS</td>
<td>JJ</td>
<td>Not NN nor NNS</td>
</tr>
<tr>
<td>JJ</td>
<td>JJ</td>
<td>Not NN or NNS</td>
</tr>
<tr>
<td>NN or NNS</td>
<td>JJ</td>
<td>Nor NN nor NNS</td>
</tr>
<tr>
<td>RB, RBR, or RBS</td>
<td>VB, VBD, VBN, VBG</td>
<td>anything</td>
</tr>
</tbody>
</table>
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

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

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

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

  $\text{PMI}(\text{word}_1, \text{word}_2) = \log_2 \frac{P(\text{word}_1, \text{word}_2)}{P(\text{word}_1)P(\text{word}_2)}$
How to Estimate Pointwise Mutual Information

- Query search engine
  - $P(\text{word})$ estimated by $\frac{\text{hits(\text{word})}}{N}$
  - $P(\text{word}_1, \text{word}_2)$ by $\frac{\text{hits(\text{word}_1 \ \text{NEAR} \ \text{word}_2)}}{N}$

$$\text{PMI}(\text{word}_1, \text{word}_2) = \log_2 \frac{\frac{1}{N} \text{hits(\text{word}_1 \ \text{NEAR} \ \text{word}_2)}}{\frac{1}{N} \text{hits(\text{word}_1)} \frac{1}{N} \text{hits(\text{word}_2)}}$$
Does phrase appear more with “poor” or “excellent”? 

Polarity(phrase) = PMI(phrase,"excellent") – PMI(phrase,"poor")
Phrases from a thumbs-up review

<table>
<thead>
<tr>
<th>Phrase</th>
<th>POS tags</th>
<th>Polarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>online service</td>
<td>JJ NN</td>
<td>2.8</td>
</tr>
<tr>
<td>online experience</td>
<td>JJ NN</td>
<td>2.3</td>
</tr>
<tr>
<td>direct deposit</td>
<td>JJ NN</td>
<td>1.3</td>
</tr>
<tr>
<td>local branch</td>
<td>JJ NN</td>
<td>0.42</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>low fees</td>
<td>JJ NNS</td>
<td>0.33</td>
</tr>
<tr>
<td>true service</td>
<td>JJ NN</td>
<td>-0.73</td>
</tr>
<tr>
<td>other bank</td>
<td>JJ NN</td>
<td>-0.85</td>
</tr>
<tr>
<td>inconveniently located</td>
<td>JJ NN</td>
<td>-1.5</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td>0.32</td>
</tr>
</tbody>
</table>
Phrases from a thumbs-down review

<table>
<thead>
<tr>
<th>Phrase</th>
<th>POS tags</th>
<th>Polarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>direct deposits</td>
<td>JJ NNS</td>
<td>5.8</td>
</tr>
<tr>
<td>online web</td>
<td>JJ NN</td>
<td>1.9</td>
</tr>
<tr>
<td>very handy</td>
<td>RB JJ</td>
<td>1.4</td>
</tr>
<tr>
<td>virtual monopoly</td>
<td>JJ NN</td>
<td>-2.0</td>
</tr>
<tr>
<td>lesser evil</td>
<td>RBR JJ</td>
<td>-2.3</td>
</tr>
<tr>
<td>other problems</td>
<td>JJ NNS</td>
<td>-2.8</td>
</tr>
<tr>
<td>low funds</td>
<td>JJ NNS</td>
<td>-6.8</td>
</tr>
<tr>
<td>unethical practices</td>
<td>JJ NNS</td>
<td>-8.5</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td><strong>-1.2</strong></td>
</tr>
</tbody>
</table>
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


• 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
Other Sentiment Tasks
• Important for finding aspects or attributes
  • Target of sentiment

• The food was great but the service was awful
Finding aspect/attribute/target of sentiment

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

<table>
<thead>
<tr>
<th>Casino</th>
<th>casino, buffet, pool, resort, beds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children’s Barber</td>
<td>haircut, job, experience, kids</td>
</tr>
<tr>
<td>Greek Restaurant</td>
<td>food, wine, service, appetizer, lamb</td>
</tr>
<tr>
<td>Department Store</td>
<td>selection, department, sales, shop, clothing</td>
</tr>
</tbody>
</table>
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 a sentence
    • “Given this sentence, is the aspect food, décor, service, value, or NONE”
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
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