Information Extraction
Seminar – Bottom-Up Sentiment Analysis

CIS, LMU München
Winter Semester 2021-2022

Alexander Fraser, CIS
Today

• Today we will take a tangent and look at another problem in information extraction: sentiment analysis
  • Today I will cover **bottom-up** sentiment analysis
• I will leave time to make a few comments about the seminar and resolve any open issues
Sentiment Analysis

• Determine if a sentence/document expresses positive/negative/neutral sentiment towards some object
Some Applications

- **Review classification**: Is a review positive or negative toward the movie?
- **Product review mining**: What features of the ThinkPad T43 do customers like/dislike?
- **Tracking sentiments toward topics over time**: Is anger ratcheting up or cooling down?
- **Prediction (election outcomes, market trends)**: Will Romney or Obama win?
Social media

- Twitter most popular
- Short (140 characters) and very informal text
- Abbreviations, slang, spelling mistakes
- 500 million tweets per day
- Tons of applications
Level of Analysis

We can inquire about sentiment at various linguistic levels:

- Words – objective, positive, negative, neutral
- Clauses – “going out of my mind”
- Sentences – possibly multiple sentiments
- Documents
Words

• Adjectives

  – objective: red, metallic
  – positive: honest important mature large patient
  – negative: harmful hypocritical inefficient
  – subjective (but not positive or negative): curious, peculiar, odd, likely, probable
Words

– Verbs
  • positive: praise, love
  • negative: blame, criticize
  • subjective: predict

– Nouns
  • positive: pleasure, enjoyment
  • negative: pain, criticism
  • subjective: prediction, feeling

Slide from Koppel/Wilson
Clauses

• Might flip word sentiment
  – “not good at all”
  – “not all good”

• Might express sentiment not in any word
  – “convinced my watch had stopped”
  – “got up and walked out”
Sentences/Documents

• Might express multiple sentiments
  – “The acting was great but the story was a bore”

• Problem even more severe at document level
Some Special Issues

• Whose opinion?

“The US fears a spill-over”, said Xirao-Nima, a professor of foreign affairs at the Central University for Nationalities.
Some Special Issues

- Whose opinion?
- Opinion about what?
Laptop Review

• I should say that I am a normal user and this laptop satisfied all my expectations, the screen size is perfect, its very light, powerful, bright, lighter, elegant, delicate... But the only thing that I regret is the Battery life, barely 2 hours... some times less... it is too short... this laptop for a flight trip is not good companion...
Even the short battery life I can say that I am very happy with my Laptop VAIO and I consider that I did the best decision. I am sure that I did the best decision buying the SONY VAIO
Some Special Issues

- Identify expressed sentiment towards several aspects of the text
  - Different features of a laptop
- Sentiment towards a specific entity
  - Person, product, company
- Emotion Analysis
  - Identify emotions in text (love, joy, anger…)
- Sarcasm
Two Approaches to Classifying Documents

• **Bottom-Up**
  – Assign sentiment to words
  – Derive clause sentiment from word sentiment
  – Derive document sentiment from clause sentiment

• **Top-Down**
  – Get labeled documents
  – Use text categorization methods to learn models
  – Derive word/clause sentiment from models
Word Sentiment

Let’s try something simple

• Choose a few seeds with known sentiment
• Mark synonyms of good seeds: good
• Mark synonyms of bad seeds: bad
• Iterate
Word Sentiment

Let’s try something simple

• Choose a few seeds with known sentiment
• Mark synonyms of good seeds: good
• Mark synonyms of bad seeds: bad
• Iterate

Not quite.

exceptional -> unusual -> weird
Better Idea
Hatzivassiloglou & McKeown 1997

1. Build training set: label all adj. with frequency > 20; test agreement with human annotators

2. Extract all conjoined adjectives

The Homestay Experience - Cultural Kaleidoscope 2006
My host’s home was very nice and comfortable. I got to try all types of food; Malaysian, Chinese, Indonesian and I loved it all. My host’s parents were very ...
www.gardenschool.edu.my/studentportal/aec/Kaleidoscope06/experience.asp - 10k - Cached • Similar pages • Note this

PriceGrabber User Rating for Watch Your Budget - PriceGrabber.com
Reviews, Camera I purchased was very nice and a bargain. There was a problem with shipping, but was resolved quickly. Buy with confidence from this vendor ...
www.pricegrabber.com/rating_getreview.php?eid=5821 - Similar pages • Note this

Testimonials
“Everybody was very nice and service was as fast as they possibly could, ... “Staff member who helped me was very nice and easy to talk to.” ...
www.sa.psu.edu/uhs/news/testimonials.cfm - 22k - Cached • Similar pages • Note this

Naxos Villages - Naxos Town or Chora Reviews. Very nice and very ...
-Did you enjoy the trip to Naxos Town: Yes it was very nice and very scenic. -In order to get to the village were there enough signs in order to find it: It ...
3. A supervised learning algorithm builds a graph of adjectives linked by the same or different semantic orientation.
4. A **clustering algorithm** partitions the adjectives into two subsets.
Even Better Idea  Turney 2001

• Pointwise Mutual Information (Church and Hanks, 1989):

$$\text{PMI}(word_1, word_2) = \log_2 \left( \frac{p(word_1 \land word_2)}{p(word_1)p(word_2)} \right)$$
Even Better Idea Turney 2001

- Pointwise Mutual Information (Church and Hanks, 1989):
  \[
  \text{PMI}(\text{word}_1, \text{word}_2) = \log_2 \left( \frac{p(\text{word}_1 \wedge \text{word}_2)}{p(\text{word}_1) p(\text{word}_2)} \right)
  \]

- Semantic Orientation:
  \[
  \text{SO}(\text{phrase}) = \text{PMI}(\text{phrase}, "\text{excellent}") - \text{PMI}(\text{phrase}, "\text{poor}")
  \]
Even Better Idea  Turney 2001

- Pointwise Mutual Information (Church and Hanks, 1989):
  \[
  \text{PMI}(\text{word}_1, \text{word}_2) = \log_2 \left( \frac{p(\text{word}_1 \land \text{word}_2)}{p(\text{word}_1)p(\text{word}_2)} \right)
  \]

- Semantic Orientation:
  \[
  \text{SO}(\text{phrase}) = \text{PMI}(\text{phrase}, "\text{excellent}") - \text{PMI}(\text{phrase}, "\text{poor}")
  \]

- PMI-IR estimates PMI by issuing queries to a search engine

\[
\text{SO}(\text{phrase}) = \log_2 \left( \frac{\text{hits}(\text{phrase \ NEAR \ "excellent"})\text{hits("poor")}}{\text{hits}(\text{phrase \ NEAR \ "poor"})\text{hits("excellent")}} \right)
\]
Resources

These -- and related -- methods have been used to generate sentiment dictionaries

• Sentinet
• General Enquirer
• …
Bottom-Up: Words to Clauses

• Assume we know the “polarity” of a word

• Does its context flip its polarity?
Prior Polarity versus Contextual Polarity

Wilson et al 2005

- **Prior polarity**: out of context, positive or negative
  
  *beautiful* → positive
  
  *horrid* → negative

- A word may appear in a phrase that expresses a different polarity in context

  “Cheers to Timothy Whitfield for the *wonderfully horrid* visuals.”

**Contextual polarity**
Example

Philip Clap, President of the National Environment Trust, sums up well the general thrust of the reaction of environmental movements: there is no reason at all to believe that the polluters are suddenly going to become reasonable.
Example

Philip Clap, President of the National Environment Trust, sums up well the general thrust of the reaction of environmental movements: there is no reason at all to believe that the polluters are suddenly going to become reasonable.
Philip Clap, President of the National Environment Trust, sums up well the general thrust of the reaction of environmental movements: there is no reason at all to believe that the polluters are suddenly going to become reasonable.
- Word token
- Word prior polarity
- Negated
- Negated subject
- Modifies polarity
- Modified by polarity
- Conjunction polarity
- General polarity shifter
- Negative polarity shifter
- Positive polarity shifter
- **Word token**
- **Word prior polarity**
  - Negated
  - Negated subject
  - Modifies polarity
  - Modified by polarity
  - Conjunction polarity
  - General polarity shifter
  - Negative polarity shifter
  - Positive polarity shifter

Slide from Koppel/Wilson
• Word token
• Word prior polarity
• **Negated**
• **Negated subject**
• Modifies polarity
• Modified by polarity
• Conjunction polarity
• General polarity shifter
• Negative polarity shifter
• Positive polarity shifter

**Binary features:**

• **Negated**
  
  For example:
  
  – *not* good
  – *does not* look very good
  
  ◆ *not only* good but amazing

• **Negated subject**

  *No politically prudent Israeli could support either of them.*
- Word token
- Word prior polarity
- Negated
- Negated subject
- **Modifies polarity**
- **Modified by polarity**
- Conjunction polarity
- General polarity shifter
- Negative polarity shifter
- Positive polarity shifter

Slide from Koppel/Wilson
- Word token
- Word prior polarity
- Negated
- Negated subject
- Modifies polarity
- Modified by polarity
- **Conjunction polarity**
- General polarity shifter
- Negative polarity shifter
- Positive polarity shifter

- **Conjunction polarity**
  - 5 values: positive, negative, neutral, both, not mod
  - *good*: negative

*good* (pos) and *evil* (neg)

Slide from Koppel/Wilson
- Word token
- Word prior polarity
- Negated
- Negated subject
- Modifies polarity
- Modified by polarity
- Conjunction polarity
- **General polarity shifter**
  
  ```
  pose little threat
  contains little truth
  ```
- **Negative polarity shifter**
  
  ```
  lack of understanding
  ```
- **Positive polarity shifter**
  
  ```
  abate the damage
  ```
Corpus → Lexicon → Step 1: Neutral or Polar? → Step 2: Contextual Polarity?

Results 2a:

Accuracy: 65.7, 65.1, 77.2
Positive F: 65.7, 65.1, 77.2
Negative F: 46.2
Neutral F: 77.2

Legend:
- Word token
- Word + Prior Polarity
- All Features

Slide from Koppel/Wilson
Results 2b

Slide from Koppel/Wilson
Conclusion:
Bottom-up sentiment analysis

- We discussed bottom-up sentiment analysis, which critically depends on lexical information
  - Easy to understand, used in some commercial products
- We may discuss top-down later, which uses text classification approaches (including deep learning)
  - This would require as background the lectures on machine learning that are coming up
• Slide sources
  – Nearly all of the slides today are from Prof. Moshe Koppel (Bar-Ilan University)

• Further reading on traditional sentiment approaches
  – 2011 AAAI tutorial on sentiment analysis from Bing Liu (quite technical)
• Thank you for your attention!