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

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

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

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

Boolean Multinomial Naïve Bayes

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

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

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.