CS544: Sentiment Analysis

February 14, 2013

Zornitsa Kozareva
USC/ISI
Marina del Rey, CA
kozareva@isi.edu
www.isi.edu/~kozareva
Search Engines

how tall is mount everest

About 821,000 results (0.11 seconds)

which is better iPad 3 or Kindle Fire

About 412,000,000 results (0.15 seconds)

Best gadget - 29,029 mentions

Mention teameven answers

The Best Tablet: Kindle Fire or iPad 2?
www.squidoo.com › ... › Gadgets › Unique Gadgets

Kindle Fire, the new tablet developed by Amazon is here, and it is currently the ... Ipad is way better, and now that they have come out with an Ipad 3 the Fire is ...
What is sentiment analysis?

- Also known as opinion mining
- Attempts to identify the opinion/sentiment that a person may hold towards an object
- It is a finer grain analysis compared to subjectivity analysis

<table>
<thead>
<tr>
<th>Sentiment Analysis</th>
<th>Subjectivity analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Subjective</td>
</tr>
<tr>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>Objective</td>
</tr>
</tbody>
</table>

(personal feelings or beliefs, facts)
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.
Google Product Search

HP Officejet 6500A Plus e-All-in-One Color Ink-jet - Fax / copier / printer / scanner
$89 online, $100 nearby ★★★★☆ 377 reviews
September 2010 - Printer - HP - Inkjet - Office - Copier - Color - Scanner - Fax - 250 sheets

Reviews

Summary - Based on 377 reviews

<table>
<thead>
<tr>
<th>1 star</th>
<th>2</th>
<th>3</th>
<th>4 stars</th>
<th>5 stars</th>
</tr>
</thead>
</table>

What people are saying

ease of use
value
setup
customer service
size
mode
colors

"This was very easy to setup to four computers."
"Appreciate good quality at a fair price."
"Overall pretty easy setup."
"I DO like honest tech support people."
"Pretty Paper weight."
"Photos were fair on the high quality mode."
"Full color prints came out with great quality."
HP Officejet 6500A E710N Multifunction Printer

Product summary  Find best price  Customer reviews  Specifications  Related items

$121.53 - $242.39 (14 stores)

Average rating ★★★★★ (144) Most mentioned
Performance (55)
Ease of Use (54)
Print Speed (10)
Connectivity (6)
More ▼

Show reviews by source
Best Buy (140)
CNET (5)
Amazon.com (3)
Twitter Sentiment

Bollen et al. (2011)

- CALM predicts DJIA 3 days later
- At least one current hedge fund uses this algorithm
Why Sentiment Analysis?

- **Product review mining:**
  What are the features of the iPad3 that the customers like?

- **Review classification:**
  Is a review positive or negative toward the movie?

- **Tracking sentiments toward topics over time:**
  Is my wife still angry?

- **Prediction (election outcomes, market trends):**
  Obama or Romney win?

- **Sentiment based text summarization**
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*
subjective vs objective

positive vs negative vs neutral

happiness vs joy vs anger vs sadness vs fear vs disgust

subjectivity analysis

sentiment analysis

emotion analysis
Sentiment Analysis

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

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 Tasks

• Simplest task:
  – Is the attitude of this text positive or negative?

• More complex:
  – Rank the attitude of this text from 1 to 5, or from -3 to +3

• Advanced:
  – Detect the target, source, or complex attitude types
Example

• A review:

(1) I bought an iPhone a few days ago. (2) **It was such a nice phone.** (3) **The touch screen was really cool.** (4) **The voice quality was clear too.** (5) Although **the battery life was not long**, that is ok for me. (6) However, **my mother was mad with me** as I did not tell her before I bought it. (7) She also thought **the phone was too expensive**, and wanted me to return it to the shop. . . .
• **Object**: an entity that can be a product, service, individual, organization, event, or topic.
  – Ex. iPhone

• **Attribute**: an object usually has two types of attributes
  – components: battery, keypad/touch screen
  – properties: size, weight, color, voice quality

• **Explicit and implicit attributes:**
  – explicit appear in the opinion:
    “the **battery life** of this phone is too short”
  – implicit do not appear in the opinion:
    “this phone is too large” (on attribute size)
• **Opinion holder**: the person or organization that expresses the opinion

• **Opinion orientation (polarity)**: positive, negative, or neutral

• **Opinion strength**: level/scale/intensity of opinion indicating how strong it is

  Ex. contented->happy->joyous->ecstatic
Target Sentiment on Twitter

- **Twitter Sentiment App**
  - Alec Go, Richa Bhayani, Lei Huang. 2009. Twitter Sentiment Classification using Distant Supervision

[Image of sentiment analysis chart]

Type in a word and we'll highlight the good and the bad

"united airlines"  Search  Save this search

Sentiment analysis for "united airlines"

- **Sentiment by Percent**
  - Negative (58%)
  - Positive (32%)

- **Sentiment by Count**
  - Positive (11)
  - Negative (23)

http://www.sentiment140.com/

- **jacobian**: OMG... Could @United airlines have worse customer service? W8g now 15 minutes on hold 4
- **12345clumsy6789**: I hate United Airlines Ceiling!!! Fukn impossible to get my conduit in this damn mess!
- **EMLandPRGbelsiu**: EML/PRG fly with Q8 united airlines and 24seven to an exotic destination. http://t.co/
- **CountAdam**: FANTASTIC customer service from United Airlines at XNA today. Is tweet more, but cell pho
Available Resources

• **Lexicons**
  • General Inquirer (Stone et al., 1966)
  • OpinionFinder lexicon (Wiebe & Riloff, 2005)
  • SentiWordNet (Esuli & Sebastiani, 2006)

• **Annotated corpora**
  • Used in statistical approaches (Hu & Liu 2004, Pang & Lee 2004)
  • MPQA corpus (Wiebe et. al, 2005)

• **Tools**
  • Algorithm based on minimum cuts (Pang & Lee, 2004)
  • OpinionFinder (Wiebe et. al, 2005)
BASELINE ALGORITHMS
Simplest Method

• **Opinion/polarity words**: dominating indicators of sentiments, especially adjectives, adverbs, and verbs, e.g. “I absolutely love this camera. It is amazing!”

• **Pre-defined opinion words**: good, terrible

• **Assign orientation score (+1, -1) to all words:**
  – positive opinion words (+1): great, amazing, love
  – negative opinion words (-1): horrible, hate
  – strength value [0, 1] can be used too

• The orientation score of paragraph/document is the sum of orientation scores of all opinion words found
Sentiment Classification in Movie Reviews


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

• Data: Polarity Data 2.0:
IMDB Data - Pang & Lee Database

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 (adapted from Pang and Lee)

- Tokenization
- Feature Extraction
- Classification using different classifiers
  - Naïve Bayes
  - MaxEnt
  - SVM
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 turns out to work better, at least on this data
Negation


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
Binarized (Boolean feature) Multinomial Naïve Bayes

• 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