Administrivia

• Hope everyone submitted the optional homework 7!
Python on Windows

Four versions present here:
1. Python command line (2.7.x)
2. IDLE Python (2.7.x)
3. Python 3.6
4. IDLE Python 3.6
Sentiment Analysis


I had a wonderful experience! The rooms were wonderful and the staff was helpful.

**Analyzed text**

- **Languages**: English (confidence: 100%)
- **Key Phrases**: wonderful experience, staff, rooms
- **Sentiment**: 96%

I had a terrible time at the hotel. The staff was rude and the food was awful.

**Analyzed text**

- **Languages**: English (confidence: 100%)
- **Key Phrases**: food, terrible time, hotel, staff
- **Sentiment**: 1%

Range [0,1]
1 = max positive
0.5 = neutral
0 = max negative
Sentiment Analysis

• *Nagender Parimi, Software Engineer at Microsoft.*
  • @Struggling – regarding score thresholds: scores less than 0.4 can be considered as having negative sentiment, while scores above 0.65 would be positive. I suggest playing with different thresholds on your specific dataset, as the thresholds can vary from one domain to another. The thresholds I gave above worked well for the datasets we evaluated.
  • As for the accuracy of the scores: we believe the service does well, but it is not perfect :-).
  • There are a few areas where we plan to improve the service, e.g. dealing with sarcasm/irony, identifying sentences that state facts vs those that express sentiment and better handling of sentences having both positive and negative sentiment.
Sentiment Analysis


The Machine Learning Approach

A more robust approach is to train models that detect sentiment. Here is how the training process works – we obtained a large dataset of text records that was already labeled with sentiment for each record. The first step is to tokenize the input text into individual words, then apply stemming. Next we constructed features from these words; these features are used to train a classifier. Upon completion of the training process, the classifier can be used to predict the sentiment of any new piece of text. It is important to construct meaningful features for the classifier, and our list of features includes several from state-of-the-art research:
Sentiment Analysis

- **N-grams** denote all occurrences of $n$ consecutive words in the input text. The precise value of $n$ may vary across scenarios, but it's common to pick $n=2$ or $n=3$. With $n=2$, for the text "the quick brown fox", the following n-grams would be generated – [“the quick”, “quick brown”, “brown fox”]

- **Part-of-speech tagging** is the process of assigning a part-of-speech to each word in the input text. We also compute features based on the presence of emoticons, punctuation and letter case (upper or lower)

- **Word embeddings** are a recent development in natural language processing, where words or phrases that are syntactically similar are mapped closer together, e.g. in such a mapping, the term *cat* would be mapped closer to the term *dog*, than to the term *car*, since both dogs and cats are animals. Neural networks are a popular choice for constructing such a mapping. For sentiment analysis, we employ neural networks that encode the associated sentiment information as well. The layers of the neural network are then used as features for the classifier.
Rotten Tomatoes

• Movie Review website [https://www.rottentomatoes.com](https://www.rottentomatoes.com)

<table>
<thead>
<tr>
<th>TOP BOX OFFICE</th>
<th>Get Tickets</th>
</tr>
</thead>
<tbody>
<tr>
<td>🍅 97% Black Panther</td>
<td>$26.8M</td>
</tr>
<tr>
<td>🍅 51% Tomb Raider</td>
<td>$23.6M</td>
</tr>
<tr>
<td>🍅 67% I Can Only Imagine</td>
<td>$17.1M</td>
</tr>
<tr>
<td>🍅 40% A Wrinkle in Time</td>
<td>$16.4M</td>
</tr>
<tr>
<td>🍅 92% Love, Simon</td>
<td>$11.8M</td>
</tr>
<tr>
<td>🍅 82% Game Night</td>
<td>$5.7M</td>
</tr>
<tr>
<td>🍅 59% Peter Rabbit</td>
<td>$5.2M</td>
</tr>
<tr>
<td>🍅 38% The Strangers: Prey At Night</td>
<td>$4.8M</td>
</tr>
<tr>
<td>🍅 47% Red Sparrow</td>
<td>$4.5M</td>
</tr>
<tr>
<td>🍅 18% Death Wish</td>
<td>$3.5M</td>
</tr>
</tbody>
</table>

View All
Critics Consensus: *Death Wish* is little more than a rote retelling that lacks the grit and conviction of the original -- and also suffers from spectacularly bad timing.
• *Death Wish* is little more than a rote retelling that lacks the grit and conviction of the original -- and also suffers from spectacularly bad timing.
Critics Consensus: *Black Panther* elevates superhero cinema to thrilling new heights while telling one of the MCU's most absorbing stories -- and introducing some of its most fully realized characters.
Sentiment Analysis

• *Black Panther* elevates superhero cinema to thrilling new heights while telling one of the MCU's most absorbing stories -- and introducing some of its most fully realized characters.
Critics Consensus:  *Red Sparrow* aims for smart, sexy spy thriller territory, but Jennifer Lawrence's committed performance isn't enough to compensate for thin characters and a convoluted story.
Sentiment Analysis

• *Red Sparrow* aims for smart, sexy spy thriller territory, but Jennifer Lawrence's committed performance isn't enough to compensate for thin characters and a convoluted story.
Sentiment Analysis

• **Mixed:** I had a wonderful experience! The rooms were wonderful and the staff was helpful. I had a terrible time at the hotel. The staff was rude and the food was awful.

*OTT ANALYZED TEXT:*

- **LANGUAGES:** English (confidence: 100%)
- **KEY PHRASES:** staff, wonderful experience, food, terrible time, hotel, rooms
- **SENTIMENT:** 7%
Sentiment Analysis

• “(Revenge of the Sith) marks a distinct improvement on the last two episodes, The Phantom Menace and Attack of the Clones ... but only in the same way that dying from natural causes is preferable to crucifixion.”

• *Star Wars: Revenge of the Sith*, Anthony Lane, The New Yorker
Sentiment Analysis

• **Objective**: Sentiment analysis refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information.

https://en.wikipedia.org/wiki/Sentiment_analysis
Sentiment Analysis

• Examples from Wikipedia:
  • The movie is surprising with plenty of unsettling plot twists.
  • (Negative term used in a positive sense in certain domains)
Sentiment Analysis

• Examples from Wikipedia:
  • Disliking watercraft is not really my thing.
  • (Negation, inverted word order)

<table>
<thead>
<tr>
<th>Analyzed text</th>
<th>JSON</th>
</tr>
</thead>
<tbody>
<tr>
<td>LANGUAGES:</td>
<td>English (confidence: 100 %)</td>
</tr>
<tr>
<td>KEY PHRASES:</td>
<td>Disliking watercraft, thing</td>
</tr>
<tr>
<td>SENTIMENT:</td>
<td>2 %</td>
</tr>
</tbody>
</table>
Sentiment Analysis

- Examples from Wikipedia:
  - You should see their decadent dessert menu.
  - (Attitudinal term has shifted polarity recently in certain domains)