



Topic 9 Sentiment Analysis

Sentiment

- Computational identification, categorization, quantification, extraction, and study of attitudes, emotional states, and/or subjective information



Figure 4-2. A review that praises some aspects and criticizes few

Different Tasks

Simplest task:

- Is the attitude of this text positive or negative?

More complex:

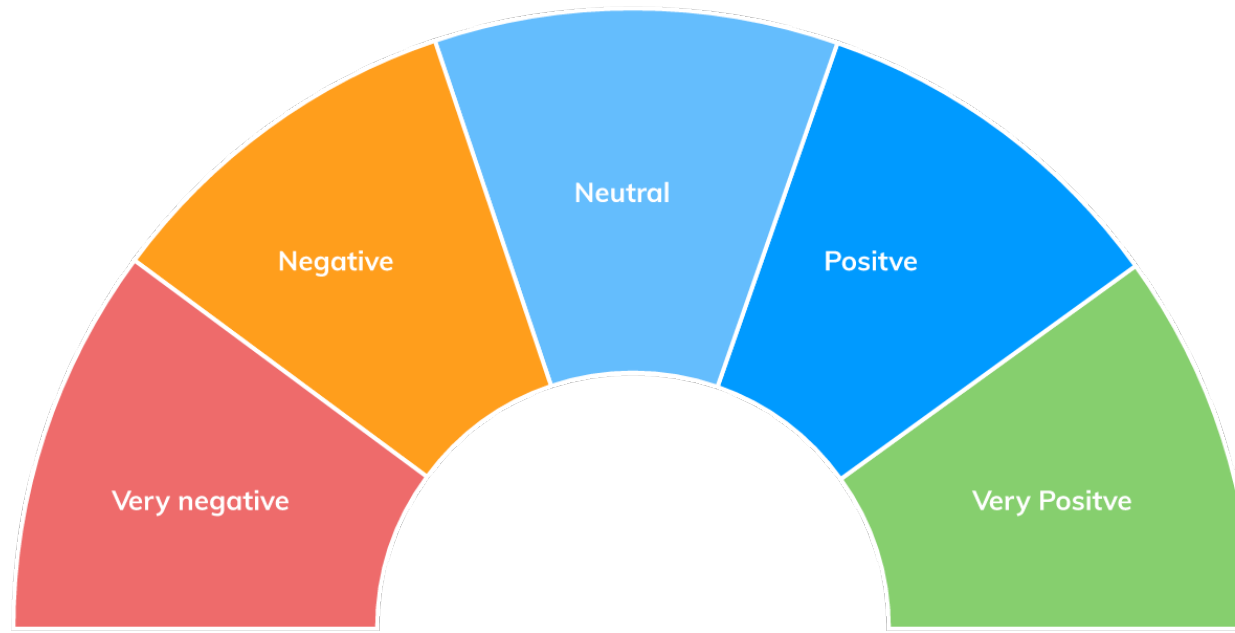
- Rank the attitude of this text from 1 to 5

Advanced:

- Detect the target, source, or complex attitude types

Polarity and Intensity

- Polarity refers to the overall sentiment conveyed by a particular text, phrase or word
- Intensity defines how strong or weak the sentiment/emotion is with respect to the context
 - Difference between “this restaurant was good” and “this restaurant was the best”



Subjectivity

- Texts can be objective or subjective
- Subjective sentence expresses some personal feelings, views, beliefs, opinions, allegations, desires, beliefs, suspicions, and speculations whereas Objective sentences are factual

Objective

- 1.The sky is blue.
- 2.The temperature outside is 20 degrees Celsius.
- 3.The Earth orbits around the Sun.

Subjective

1. The movie I watched last night was incredibly boring.
2. I love listening to music while driving.
3. The beach is the most relaxing place in the world.

How are these calculated?

- AFINN Lexicon
- SentiWordNet
- Pattern
- VADER (Valence Aware Dictionary and Sentiment Reasoner)
- Multi-Perspective Question Answering

Lexicon-based solutions

- “Lexicons” or lists of positive and negative words are created
- Can use a series of heuristics to encode how contextual elements increment, decrement, or negate the sentiment of text

Advantages

- Sorting Data at Scale
- Real-Time Analysis
- Consistent criteria

ML and DL solutions

- Sentiment Analysis is just another type of classification problem
- Common models used for sentiment analysis:
 - Naïve Bayes
 - SVM
 - Random Forest
 - LSTMs
 - Transformers

Example dataset features

- Counts of number of positive and negative words
- POS tags
 - adjectives and adverbs may be more likely to indicate sentiment than nouns or verbs
- Word occurrence may matter more than word frequency
 - The occurrence of the word “fantastic” may tell a lot
 - If it occurs 5 times, it may not tell much more
- Boolean Multinomial Naïve Bayes
 - Clip all the word counts in each document at 1

Example using TextBlob

- Sentiment calculation from the pattern library
- Measure polarity and subjectivity
- It can be used in place for NLTK and spaCy library

```
doc=TextBlob("We are having fun here")
```

```
doc.polarity # -1 to +1
```

```
doc.subjectivity # 0 to +1
```

```
doc.sentiment # returns both
```

Example using TextBlob

- Ignores one-letter words
- Ignore words it doesn't know anything about
- Finds words and phrases it can assign polarity and subjectivity to, and it averages them all together for longer text

Example using Vader

- A lexical dictionary contains regular words, as well as phrases like "not good", "really cool", or "the best", emoticons like happy faces :) and sad faces :(, and abbreviations that convey strong feelings such as LOL or OMG
- Lexical features were rated for the polarity and intensity on a scale from “-4: Extremely Negative” to “+4 Extremely Positive” by 10 independent human raters
- The average of these scores is used as the sentiment indicator for each lexical feature in the dictionary

Example using Vader

- The compound score is the sum of positive, negative & neutral scores which is then normalized between -1 (most extreme negative) and +1 (most extreme positive).

$$\frac{x}{\sqrt{x^2 + \alpha}}$$

x is the sum of the sentiment scores of the constituent words

alpha is a normalization parameter that we set to 15

- Heuristics for capitalization, punctuation, sentences containing “but”, degree modifiers, negation

Example using Vader

```
import nltk
nltk.download('vader_lexicon')
from nltk.sentiment.vader import SentimentIntensityAnalyzer

sid = SentimentIntensityAnalyzer()

scores = sid.polarity_scores(text)
print(scores)
# {'neg': 0.0, 'neu': 0.0, 'pos': 0.0, 'compound': 0.0}
```

Challenges

- Negation, inverted word ordering
- Metaphors
- Sarcasm
- Negative term used in a positive way
- Context
- Comparison
- Subtlety

Aspect-level Sentiment Analysis

- If you identify the sentiment of a piece of text it may be beneficial to know what is being written about
- Supervised or unsupervised

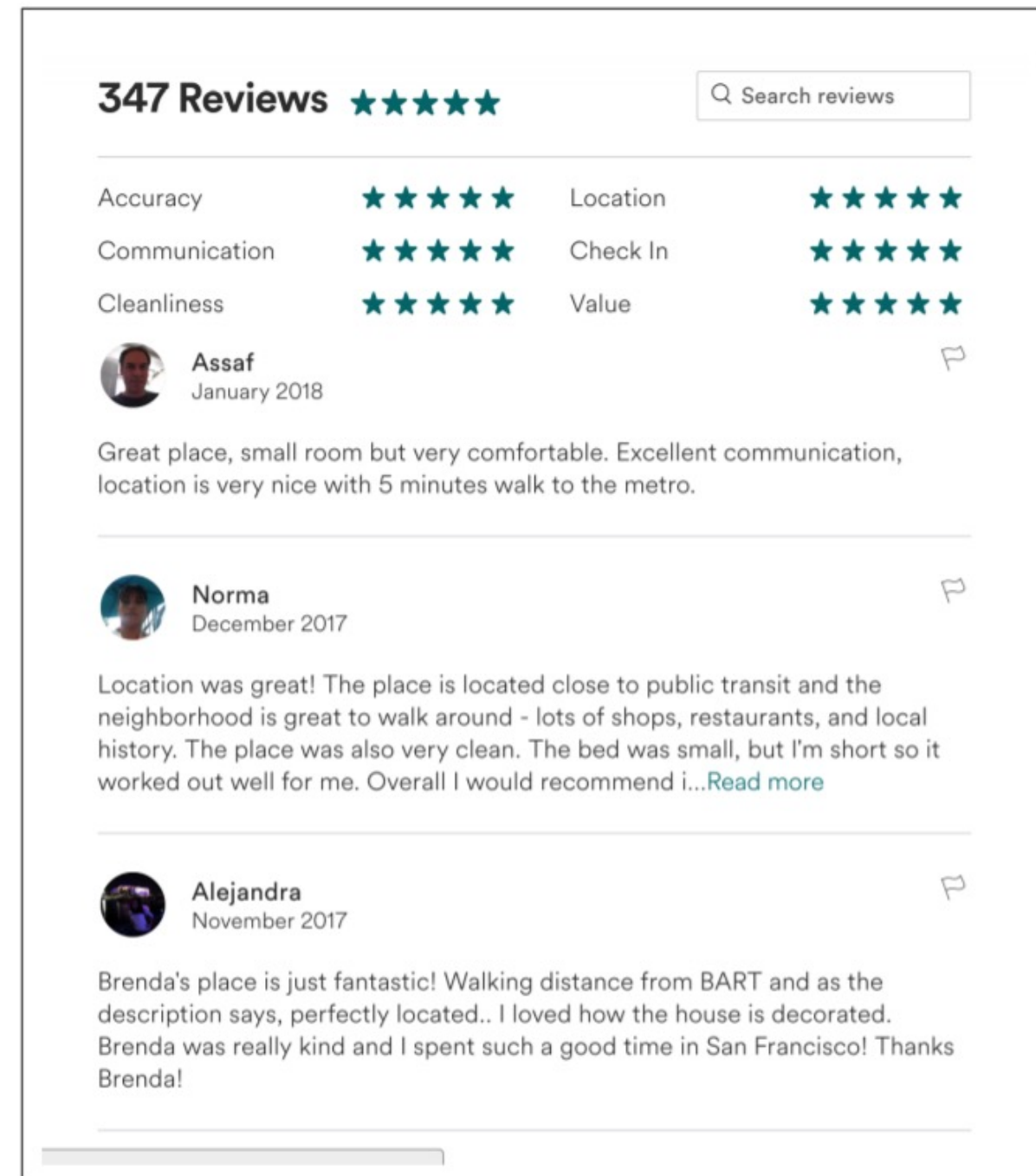


Figure 9-15. Aspect-level ratings on reviews given in a travel website

Next time

- Information Extraction