Sentiment Analysis of Summaries

Eric Allen

Summaries

- There are now online reviews of a wide variety of things on many websites across the internet

- Traditional sources, but also influencers on social media, etc.

- *“I bought this television a year ago as an anniversary gift. My whole family loves it! The picture quality is fantastic. It’s a bit pricey, but worth every penny. And I have come to love the color, although avocado green would not have been my first choice.”*

Selecting salient summaries

- *“I bought this television a year ago as an anniversary gift. My whole family loves it! The picture quality is fantastic. It’s a bit pricey, but worth every penny. And I have come to love the color, although avocado green would not have been my first choice.”*

How to select a summary?

- *“I bought this television a year ago as an anniversary gift. My whole family loves it! The picture quality is fantastic. It’s a bit pricey, but worth every penny. And I have come to love the color, although avocado green would not have been my first choice.”*

How to select a summary?

- One solution: Human readers (e.g., Mechanical Turk)
- Advantages:
    - Context understood, inappropriate (e.g., ungrammatical or “too grammatical”) summaries easily avoided, sarcasm detectable
- Disadvantages:
    - Time intensive, subjective, costly

Automatic summary selection

Can we use a sentiment analysis engine to automatically select summaries?

Sentiment Analysis

- Given a short block of text, automatically determine whether the text is saying positive or negative things

- Extra points for determining *what* it is saying positive or negative things about

- Double extra points for determining how positive or negative it is

Sentiment Analysis

- A simple strategy for building a sentiment analysis engine:
  - Find a large set of text samples (each preferably around the size of a product review) with labels on them mentioning whether they are saying positive or negative things (e.g., a 1-5 star rating)
  - Turn each example into a histogram measuring the number of occurrences of each 4 letter n-gram in the text:
    - “The picture” breaks into these n-grams:
    - “The ” “he p” “e pi” “ pic” “pict” “ictu” “ctur” “ture”

Sentiment Analysis

- Treat these histograms as (very) high dimensional vectors (with labels) and feed them to a machine learning algorithm

- The trained algorithm can now look at a new text sample, turn it into a histogram, and predict the star rating we might attach to that text (i.e., its sentiment)

- A similar technique is used to build email spam detectors

Autocuration Plan A

- Break all reviews into a collection of sentences
- Rate the sentences with our sentiment analysis engine
- Choose the *n* highest rated sentences

Automatic Summary Plan A

- Problem:
- Some sentences are positive sentiment but not appropriate for use in an advertisement:
  - *“My hotel room didn’t smell smoky at all.”*

Insight

- Problematic sentences tend to occur in the context of a larger, mixed review:
  - *“I walked into the hotel lobby and I could hardly breathe! The lobby was full of smoke. I quickly checked in and escaped to my room. Once there, I was quite relieved: My hotel room didn’t smell smoky at all.”*

Automatic Summary Plan B

- Run whole reviews through the sentiment analysis engine along with individual sentences

- Rank each sentence according to the *geometric mean* of the sentiment of the whole review and the sentiment of the sentence

- Geometric mean biases toward sentences that align with the sentiment of their context

Automatic Summary Plan B

- *“I bought this television a year ago as an anniversary gift. My whole family loves it! The picture quality is fantastic. It’s a bit pricey, but worth every penny. And I have come to love the color, although avocado green would not have been my first choice.”*

Additional Advantages

- Summary generation can be included in a tool, enabling humans to explore items and summaries
- Enables *machine exploration* of well-liked items by scouring reviews and measuring the sentiments of summaries