

# Studying the Impact of Multimodality in Sentiment Analysis

Ahmad Elshenawy  
Steele Carter



# Goals/Motivation

- How are judgments influenced by different modalities?
- Compare sentiment contributions of different modalities
- Use Interannotator agreement to measure objectivity of sentiment and ease of judgment
- Observe how results change for fine grained judgments of review chunks

# Background/prior work

- [Towards Multimodal Sentiment Analysis: Harvesting Opinions from the Web](#) (Morency et al)
  - Built sentiment classifiers using features from 3 different modalities:
    - Text
    - Audio
    - Video
  - Created YouTube corpus of video reviews
  - Found that integrating all 3 modalities yields best performance

# Corpus

- We created our own corpus of Youtube video reviews, consisting of 3-5 minute long book reviews.
- Originally 35 videos were found and analyzed, but the experiment uses only 20 videos.
  - corpus reduced primarily due to cost concerns
  - 6 positive, 6 negative, 8 neutral
- Originally video transcriptions were obtained via crowdsourcing
  - was way too slow, and way too expensive

# Annotation

- Transcribed each video by hand
  - Labeled disfluencies (um, er, etc.)
- Also labeled our own evaluations of sentiment for comparison and spam filtering
- Added timestamps dividing transcriptions into chunks

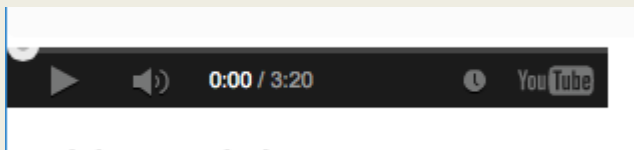
# Modalities

We experiment on four different modalities here:

- **Text only:** typical in sentiment analysis, workers are given only a piece of text.

'You wrote House of Leaves, remember.' 'Yes.' 'House of Leaves, amazing book, that simultaneously destroyed and reconstructed modern storytelling, that house of leaves?' 'mm hmmm.' 'and you think a fitting addition to your cannon, is a fairy tale, told quite unnecessarily as a long-form poem?' 'Absolutely.'

- **Audio only:** workers are given an audio-only piece of the review.



# Modalities - cont'd

- **Video only:** workers are given a video piece of the review where the video is muted, and they are given no option to increase the volume.



- **Audio/Video:** a complete piece of a video, with sound and video intact.

# Video Chunks

- Videos were annotated with timestamps, breaking up videos into ~20-30 second chunks, typically also demarcating new topics within the review.
- A HIT was designed where workers are presented with 5 of these chunks, and asked to judge the sentiment of that chunk.



# HIT Design

- Experiment ended up needing 8 Mechanical Turk HITs.
  - One set of HITs for each modality.
    - Text only, audio only, video only, audio/video
  - One set of HITs for chunks vs whole reviews
- Required **a lot** of javascript and HTML coding
- Collected 10 judgments per video/fragment, paying about \$0.15 per task.
  - 20 video HITs per modality
  - 21 5-chunk HITs per modality

## Preview HITS

1 Select HIT Template 2 Upload Input Data 3 Preview 4 Confirm and Publish

This is how your HIT will look to Workers. Make sure that any variables in the HIT are correctly replaced by your input data, then click "Next".

### Analyze the sentiment in text fragments of a review

Analyze the sentiment in text fragments of a review

**Requester:** Ahmad Elshenawy **Reward:** \$0.15 per HIT **HITs available:** 21 **Duration:** 1 Hours

**Qualifications Required:** HIT Approval Rate (%) for all Requesters' HITs greater than or equal to 95 ,  
Number of HITs Approved greater than or equal to 500

#### HIT Preview

##### Instructions

- Instruction #1: First please fill out a brief survey, providing information about gender, age group and country,
- Instruction #2: Please read five fragments from a review, and tell us whether you feel that fragment was positive, negative or neutral in tone in accordance with the table immediately below.

5	Strongly Positive	Select this if the item embodies emotion that was extremely happy or excited toward the topic. For example, "Their customer service is the best that I've seen!!!!"
4	Positive	Select this if the item embodies emotion that was generally happy or satisfied, but the emotion wasn't extreme. For example, "Sure I'll shop there again."
3	Neutral	Select this if the item does not embody much of positive or negative emotion toward the topic. For example, "Yeah, I guess it's ok." or "Is their customer service open 24x7?"
2	Negative	Select this if the item embodies emotion that is perceived to be angry or upsetting toward the topic, but not to the extreme. For example, "I don't know if I'll shop there again because I don't trust them."
1	Strongly Negative	Select this if the item embodies negative emotion toward the topic that can be perceived as extreme. For example, "These guys are terrific... NOTTTTT!!!!!!" or "I will NEVER shop there again!!!"

## Instructions

Number of HITs Approved greater than or equal to 500

### HIT Preview

4	Positive	Select this if the item embodies emotion that was generally happy or satisfied, but the emotion wasn't extreme. For example, "Sure I'll shop there again."
3	Neutral	Select this if the item does not embody much of positive or negative emotion toward the topic. For example, "Yeah, I guess it's ok." or "Is their customer service open 24x7?"
2	Negative	Select this if the item embodies emotion that is perceived to be angry or upsetting toward the topic, but not to the extreme. For example, "I don't know if I'll shop there again because I don't trust them."
1	Strongly Negative	Select this if the item embodies negative emotion toward the topic that can be perceived as extreme. For example, "These guys are terrific... NOTTTTT!!!!!" or "I will NEVER shop there again!!!"

#### Pre-Survey

**1. What is your gender?**

- Male
- Female

**2. What is your age group?**

- select one -

**3. Country of Residence**

### Fragment 1

Fragment 1: 169seconds - 197seconds



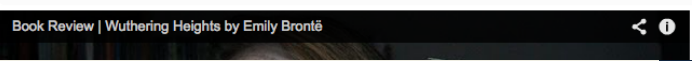
**Fragment 1 Partial Transcription**  
Please transcribe the first 10 words of this fragment. Include utterances like 'umm', 'like' and so on:

**Fragment 1 Sentiment**

- select one -

### Fragment 2

Fragment 2: 140seconds - 155seconds



Example of an Audio/Video Chunk HIT

Qualifications Required: HIT Approval Rate (%) for all Requesters' HITs greater than or equal to 95 ,  
Number of HITs Approved greater than or equal to 500

### HIT Preview

#### Fragment 1

There's the bit where the pastor has to deal with all the crazy ladies in his congregation and one woman didn't show up to the meeting and all the other ladies are pissed that she isn't there, and then the pastor says, 'Isn't it funny that the rapture finally happens and the only person to be taken away is Cynthia?' This is the same lady again who says that, 'She once heard that the best thing for the planet would be for everyone to stay in one place for five years. No more transience, no more geographical cures, no more petro holidays. Just a simple commitment to one spot.'

#### Fragment 1 Sentiment

- select one -

#### Fragment 2

I had to take it segmented. I had to read a chapter and then I watched clueless, and then I read a chapter, and then I watched Romy and Michele's High School Reunion, and then I read a chapter, and I watched Heather's -- Heather's probably wasn't the best decision in hindsight.

#### Fragment 2 Sentiment

- select one -

#### Fragment 3

Anyway there's no point in me telling you the story of Pride and Prejudice, everyone knows the story of Pride and Prejudice. You know how it ends. I'm not gonna tell you how it ends, just in case you haven't read this novel, I mean you have had 200 years, so you know, it's kinda your own fault if you read spoilers anywhere.

#### Fragment 3 Sentiment

- select one -

#### Fragment 4

Example of a Text Chunk HIT

# Spam detection/prevention

- HITs with audio, ask workers to transcribe first 10 words
- Label Gold sentiment chunks
  - Discard HITs that disagree with Gold polarity (eg if Gold is 5, discard 3 but keep 5)
  - Issue: can't label video only modality
- Compare submissions to average MTurk worker judgments
- Currently, spam filtration has caught 175+ spam submissions

# Results

- In progress
- Results so far...

experiment	Audio Fragments	Audio Full	AV Fragments	AV Full	Text Fragments	Text Full	Video Fragments	Video Full
kappa	0.7704488	0.4029066	XXXXXXXX	0.3512912	0.4193037	0.3348412	0.2079012	0.1747049

# Potential Analysis

- Interannotator Agreement
- Agreement between modalities
- Compare to Gold
- Compare Chunk deviation from full video sentiment judgment



# Reference

- Morency, Louis-Phillipe and Mihalcea, Rada and Doshi, Payal. [Towards Multimodal Sentiment Analysis: Harvesting Opinions from the Web](#), Proceedings of ICMI '11 Proceedings of the 13th international conference on multimodal interfaces, p. 169-176.