What Makes People Laugh? Multimodal Humor Detection and Analysis in Videos

Zixiaofan Yang, Lin Ai, Julia Hirschberg
Columbia University

INTRODUCTION

- Properties of humor
  1. Producer + perceiver
  2. Positive emotional reactions (laughter)
  3. Highly individualistic & cultural specific

↓

Lack of multimedia data annotated with humor

- Our approach
  - Collect videos potentially with humorous utterances
  - Use time-aligned user comments to automatically generate humor labels
  - Analyze the speech, text, and visual characteristics of humorous expressions
  - Train models for predicting humor

DATA COLLECTION

- Time-aligned comments on *bilibili.com*
  - Users can post comments about a specific scene while watching the video
- 'Papi酱' - A Chinese online celebrity
  - Famous for discussing trending topics in a humorous way
  - 4 million subscribers, 296 million views
  - 100 videos, 93593 comments
- Video segmentation
  - One-second unit level: 24K segments
  - Inter-pausal unit (IPU) level: 8K segments
- Laughing indicators
  - '233' (internet meme) + 'hh' or '哈哈' (onomatopoeia of laughter)

Audiences tend to respond to humor in videos with laughing

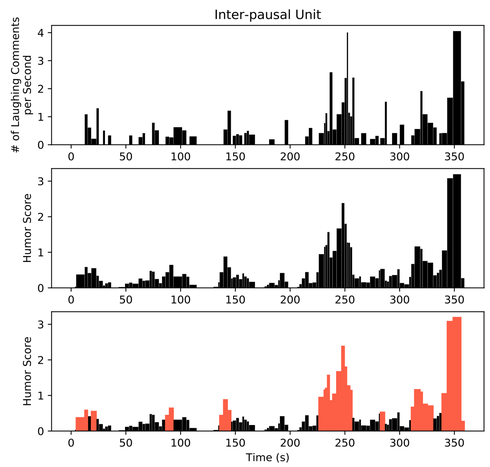
A high volume of laughing comments at a given time

↓

HUMOR

CONSTRUCTING UNSUPERVISED LABELS

- Response time estimation
  - Users typically do not pause to comment
  - Response Time = reaction time + typing time
- An example of contextual smoothing and automatic labeling on IPU level



# of laughing comments for each IPU segment before smoothing

After contextual smoothing with response time

After setting a threshold to distinguish humor from non-humor segments

- Automatic labels' accuracy on a manually annotated test set
  - One-second units: 0.78
  - Inter-pausal units: 0.76

FEATURE EXTRACTION

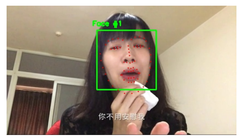
- Acoustic-Prosodic features: pitch, intensity, speaking rate, etc.
- Transcript-based lexical features: word categories such as function words, affect words, social words, etc.
- Visual features: frame similarity, body pose, facial landmarks

HUMOR ANALYSIS

- Acoustic-prosodic analysis
  - Humor techniques of exaggeration and bombast

	One-second Unit		Inter-pausal Unit (IPU)	
	t	p	t	p
Pitch existence	8.71	p<0.001	1.57	p=0.116
Pitch min	3.68	p=0.403	-2.20	p=0.028
Pitch max	4.62	p<0.001	5.52	p<0.001
Pitch mean	6.21	p<0.001	4.37	p<0.001
Pitch range	2.40	p=0.016	6.55	p<0.001
Pitch stddev	0.93	p=0.352	3.64	p<0.001
Intensity min	6.91	p<0.001	4.22	p<0.001
Intensity max	16.88	p<0.001	11.76	p<0.001
Intensity mean	7.02	p<0.001	3.82	p<0.001
Intensity range	-5.02	p<0.001	-3.30	p<0.001
Intensity stddev	-3.57	p<0.001	-2.68	p<0.001
Speaking rate	-10.12	p<0.001	-10.16	p<0.001

- Lexical analysis
  - Negative polarity: Negative emotion such as anxiety
  - Human-centeredness: first person pronouns
- Visual analysis
  - Movement of key points above shoulder, head-turning



EXPERIMENTS

- Training set: 70 videos with automatic humor labels
- Test set: 30 videos with manual humor labels
- F1-scores of random forest classifiers

	One-second Unit	Inter-pausal Unit (IPU)
Speech	0.71	**0.76**
Text	0.70	0.70
Visual	0.72	0.72
Speech + Text	0.72	**0.76**
Speech + Visual	**0.73**	0.75
Text + Visual	0.72	0.72
All Features	**0.73**	0.75

FUTURE WORK

- Collecting videos with different types of humor
- Other live streaming websites with time-aligned user comments