







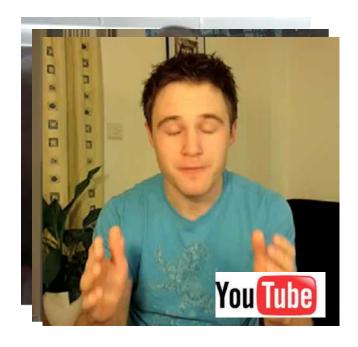
Computational Modeling of Human Multimodal Language: The MOSEI Dataset and Interpretable Dyanamic Fusion

Paul Pu Liang

Committee: Louis-Philippe Morency, Ruslan Salakhutdinov

Progress of Artificial Intelligence

Multimedia Content



Intelligent Personal Assistants





Robots and Virtual Agents



Multimodal Communicative Behaviors

Language

- > Lexicon
- > Syntax
- > Pragmatics

Acoustic

- > Prosody
- > Vocal expressions

Visual

- > Gestures
- > Body language
- > Eye contact
- > Facial expressions



Sentiment

- Positive
- Negative

Emotion

- > Anger
- Disgust
- > Fear
- > Happiness
- Sadness
- > Surprise

Personality

- Confidence
- > Persuasion
- Passion

New Dataset: MOSEI

23,000 video segments 3 modalities

Language: And he I don't think he got mad when hah I don't know maybe.

Vision:

Acoustic:



(frustrated voice)



(angry voice)

Too much too fast, I mean we basically just All I can say is he's a pretty average guy. get introduced to this character...



Contradictory

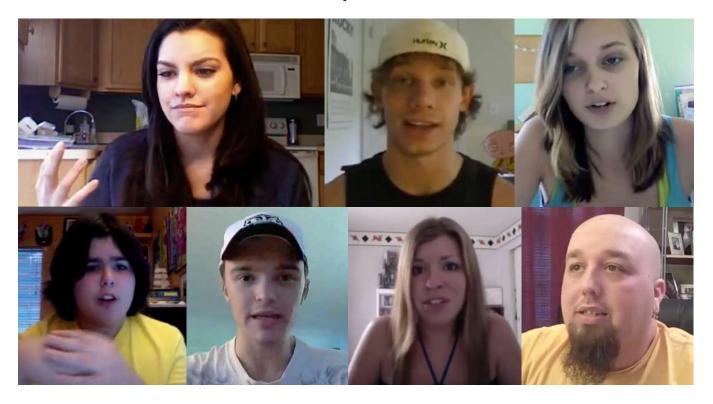




(disappointed voice)

MOSEI Dataset

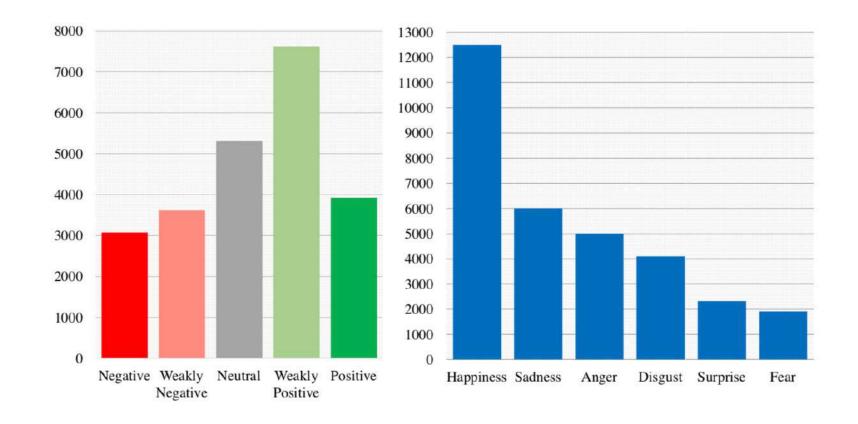
1,000 speakers



250 topics

```
consumer
                                         independent movie
                        testimony monologue application sector business counter marketplace entrepreneurship
            courses industry
advertising updates hearing consulting online advertising updates hearing consulting online pricing faq instruction
             debates thoughts economics investing hear distress
                         consumers Response
         buyers lecture listener review home narrative financing presentation seminar customers equity remarks
                     th loans updated integrated questions update products politics analysis firms retail meeting
      statement
                 speeches advertisers meeting
                                                        financial businesses
                                           product religious sociology
                                                               committee economic
           placement eulogy description employers
                                             speaker topic outsourcing
                evaluation
                                            dialogue banking seeing
                retailers
                            convention
                                          customer marketing
```

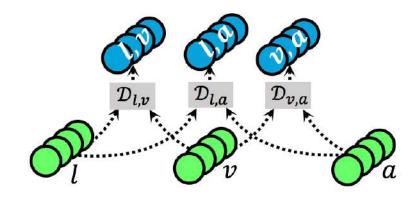
Annotation Distributions



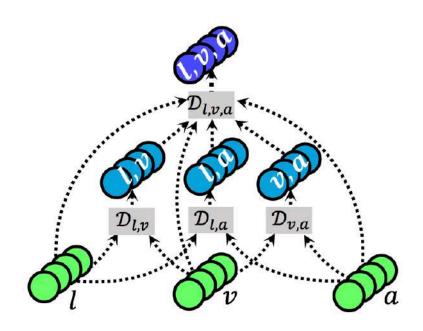






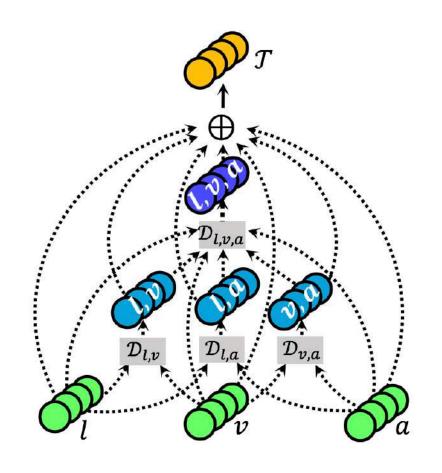


bimodal



trimodal

bimodal



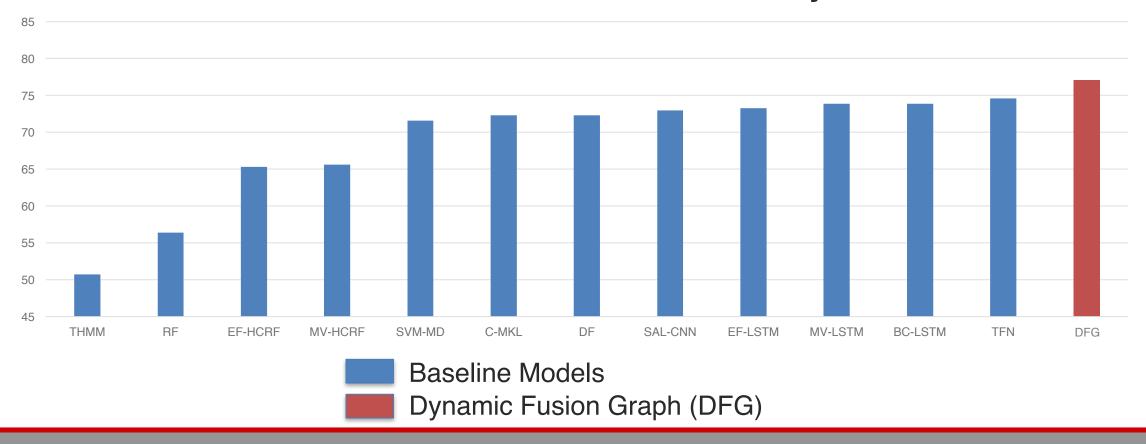
multimodal representation

trimodal

bimodal

State-of-the-art Results

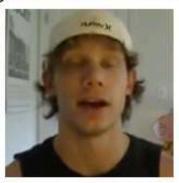
MOSEI Multimodal Sentiment Analysis



Interpretable Fusion

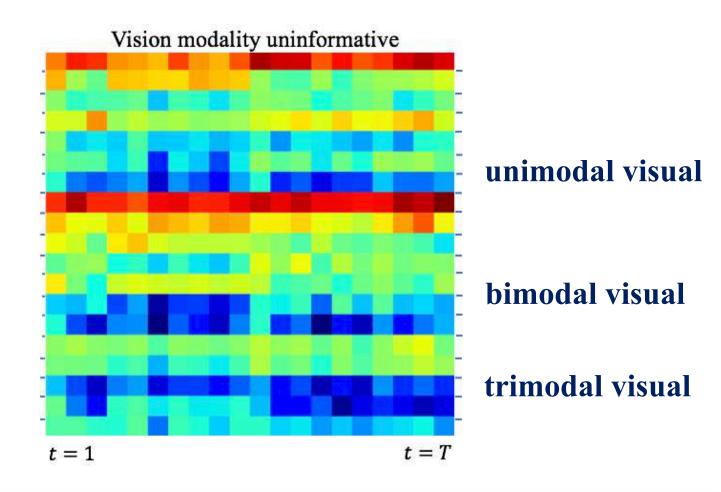
Too much too fast, I mean we basically just get introduced to this character...

Uninformative





(angry voice)



Computational Modeling of Multimodal Language

MOSEI Dataset Diversity in samples, topics, speakers and annotations **Dynamic Fusion Graph** State-of-the-art results Interpretable fusion

The End!

Email: pliang@cs.cmu.edu

Workshop @ ACL 2018

First Workshop and Grand Challenge
on Computational Modeling of Human Multimodal Language

multicomp.cs.cmu.edu/acl2018multimodalchallenge/