Results of First Grand-Challenge and Workshop on Human Multimodal Language (Challenge-HML)

Presented by

Zhun Liu

CMU-MOSEI Dataset

23,000 video segments
3 modalities

CMU-MOSEI Dataset

1,000 speakers
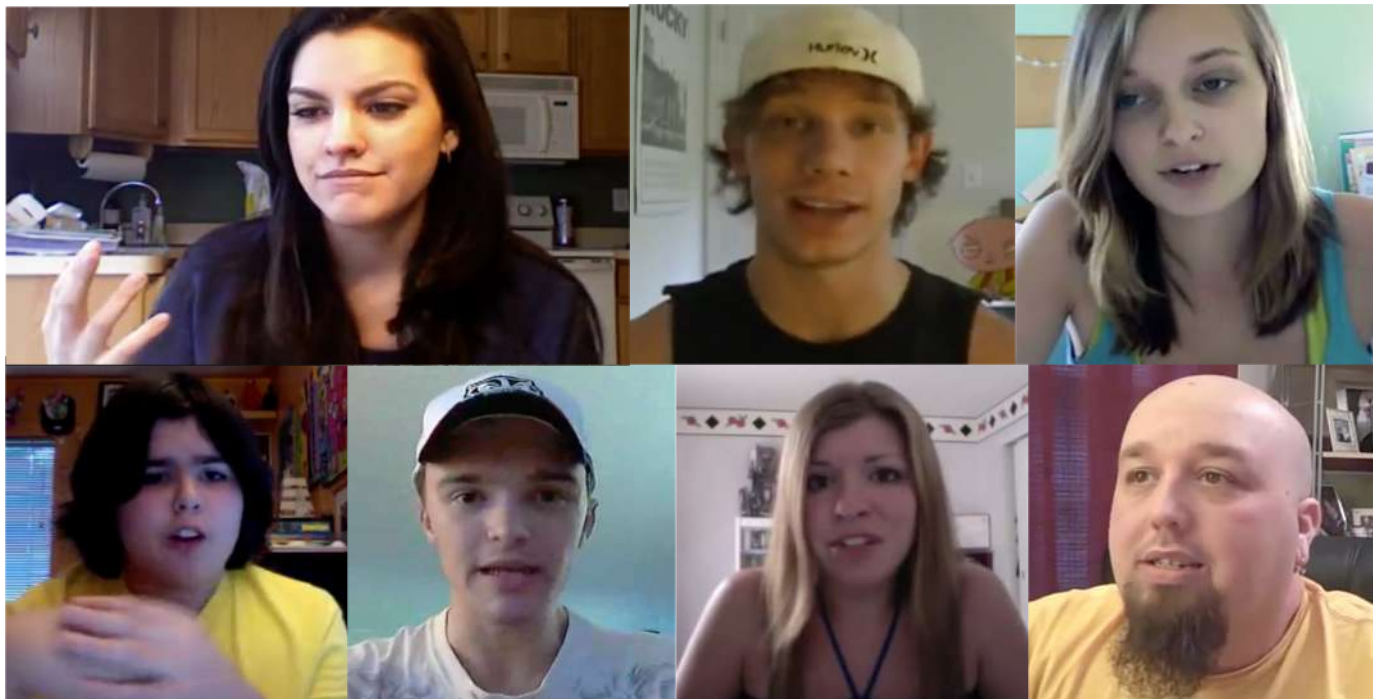
250 topics

CMU-MOSEI Dataset

Dataset	# S	# Sp	Mod	Sent	Emo	TL (hh:mm:ss)
CMU-MOSEI	**23,453**	**1,000**	$\{l, v, a\}$	✓	✓	**65:53:36**
CMU-MOSI	2,199	98	$\{l, v, a\}$	✓	✗	02:36:17
ICT-MMMO	340	200	$\{l, v, a\}$	✓	✗	13:58:29
YouTube	300	50	$\{l, v, a\}$	✓	✗	00:29:41
MOUD	400	101	$\{l, v, a\}$	✓	✗	00:59:00
SST	11,855	–	$\{l\}$	✓	✗	–
Cornell	2,000	–	$\{l\}$	✓	✗	–
Large Movie	25,000	–	$\{l\}$	✓	✗	–
STS	5,513	–	$\{l\}$	✓	✗	–
IEMOCAP	10,000	10	$\{l, v, a\}$	✗	✓	11:28:12
SAL	23	4	$\{v, a\}$	✗	✓	11:00:00
VAM	499	20	$\{v, a\}$	✗	✓	12:00:00
VAM-faces	1,867	20	$\{v\}$	✗	✓	–
HUMAINE	50	4	$\{v, a\}$	✗	✓	04:11:00
RECOLA	46	46	$\{v, a\}$	✗	✓	03:50:00
SEWA	538	408	$\{v, a\}$	✗	✓	04:39:00
SEMAINE	80	20	$\{v, a\}$	✗	✓	06:30:00
AFEW	1,645	330	$\{v, a\}$	✗	✓	02:28:03
AM-FED	242	242	$\{v\}$	✗	✓	03:20:25
Mimicry	48	48	$\{v, a\}$	✗	✓	11:00:00
AFEW-VA	600	240	$\{v, a\}$	✗	✓	00:40:00
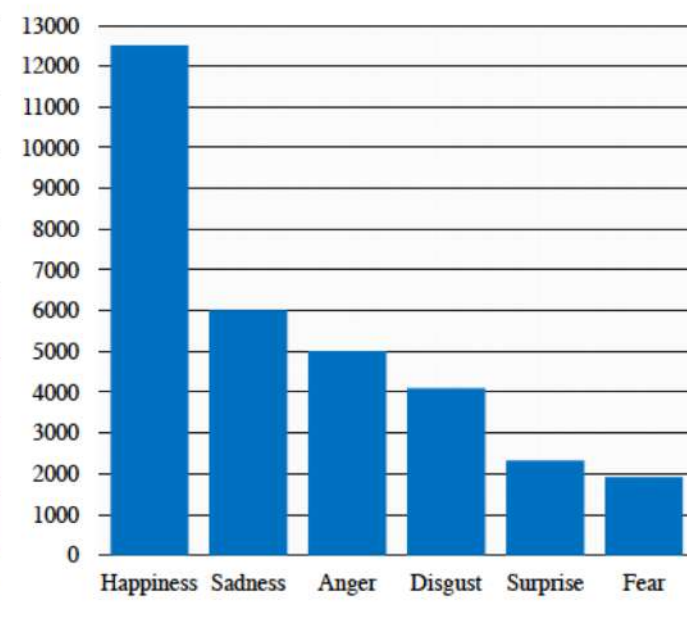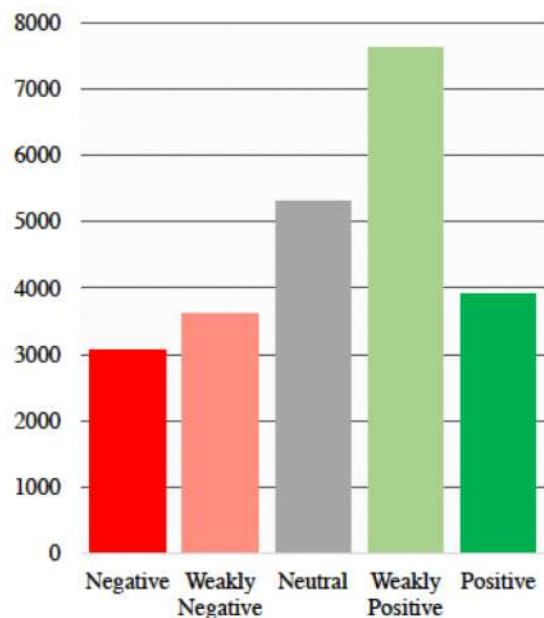
CMU-MOSEI: Tasks and Evaluation Metrics

▪ Sentiment Analysis:
- Sentiment Intensity Regression: Mean Absolute/Square Error;
- Binary Sentiment Classification: Accuracy, precision, recall, F1-score;
- Fine-grained (7-class) Sentiment Classification: Same as above, in 7 classes.

▪ Emotion Recognition:
- Multilabel Emotion Intensity Regression: Average/Weighted MAE/MSE;
- Multilabel Binary Emotion Classification: Average/Weighted Accuracy, P/R, F1;
- Multilabel Fine-grained (4-class) Emotion Classification: Same as above, in 4 classes.

CMU-MOSEI: Tasks and Evaluation Metrics

- Sentiment Analysis:
  - **Sentiment Intensity Regression: Mean Absolute/Square Error;**
  - Binary Sentiment Classification: Accuracy, precision, recall, F1-score;
  - Fine-grained (7-class) Sentiment Classification: Same as above, in 7 classes.

- Emotion Recognition:
  - **Multilabel Emotion Intensity Regression: Average/Weighted MAE/MSE;**
  - Multilabel Binary Emotion Classification: Average/Weighted Accuracy, P/R, F1;
  - Multilabel Fine-grained (4-class) Emotion Classification: Same as above, in 4 classes.

Task I: Sentiment Analysis – Best Performance

1st place in HML Challenge Sentiment Analysis Track:

Team **Anticipatory Computing Lab:**

Saurav Sahay, Shachi H Kumar, Rui Xia, Jonathan Huang and Lama Nachman

Task II: Emotion Recognition – Best Performance

1ˢᵗ place in HML Challenge Emotion Recognition Track:

Team **Edinburgh-G25:**

Jennifer Williams, Steven Kleinegesse, Ramona Comanescu and Oana Radu

Thank you for joining us today!