Prediction of Personality First Impressions With Deep Bimodal LSTM

Noa Glaser Karen Yang

CS 231N Spring 2017

1. Background and Motivation

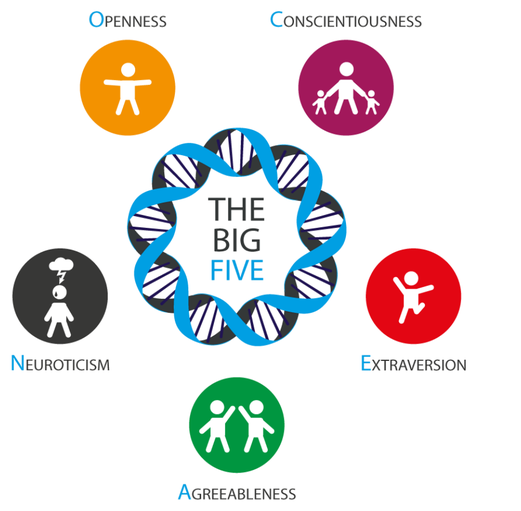

Figure 1. "Big Five" personality trait (OCEAN)

From job interviews to first dates, a first impression can make or break an interaction. Human form judgments in the first 100ms of interaction.[7] Can an AI predict apparent personality traits given a short video?

We propose a Deep Bimodal Regression LSTM model that extracts temporally ordered visual and audio features from a video clip to predict people's first impression on the "big five" personality traits widely used in psychology researches and personality profiling by hiring managers.

2. Problem Statement

Dataset

- 6,000 HD YouTube videos (duration ~15s) of people facing a webcam.
- Videos subjects vary in gender, age, nationality, and ethnicity. [8]

Evaluation Metric $\frac{1}{5N}\sum_{j=1}^{5}\sum_{i=1}^{N} 1 - \|\text{groundTruth}_{ij} - \text{predicted}_{ij}\|$

- Labeled first impressions of the five personality traits on the range [0,1]
- Crowd sourced through Amazon Mechanical Turk

Model Input and Output

Figure 2. Model Input

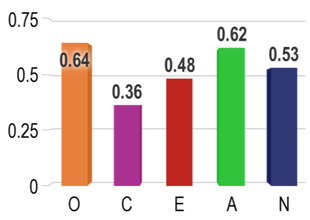

Figure 3. Model Output

3. Methods

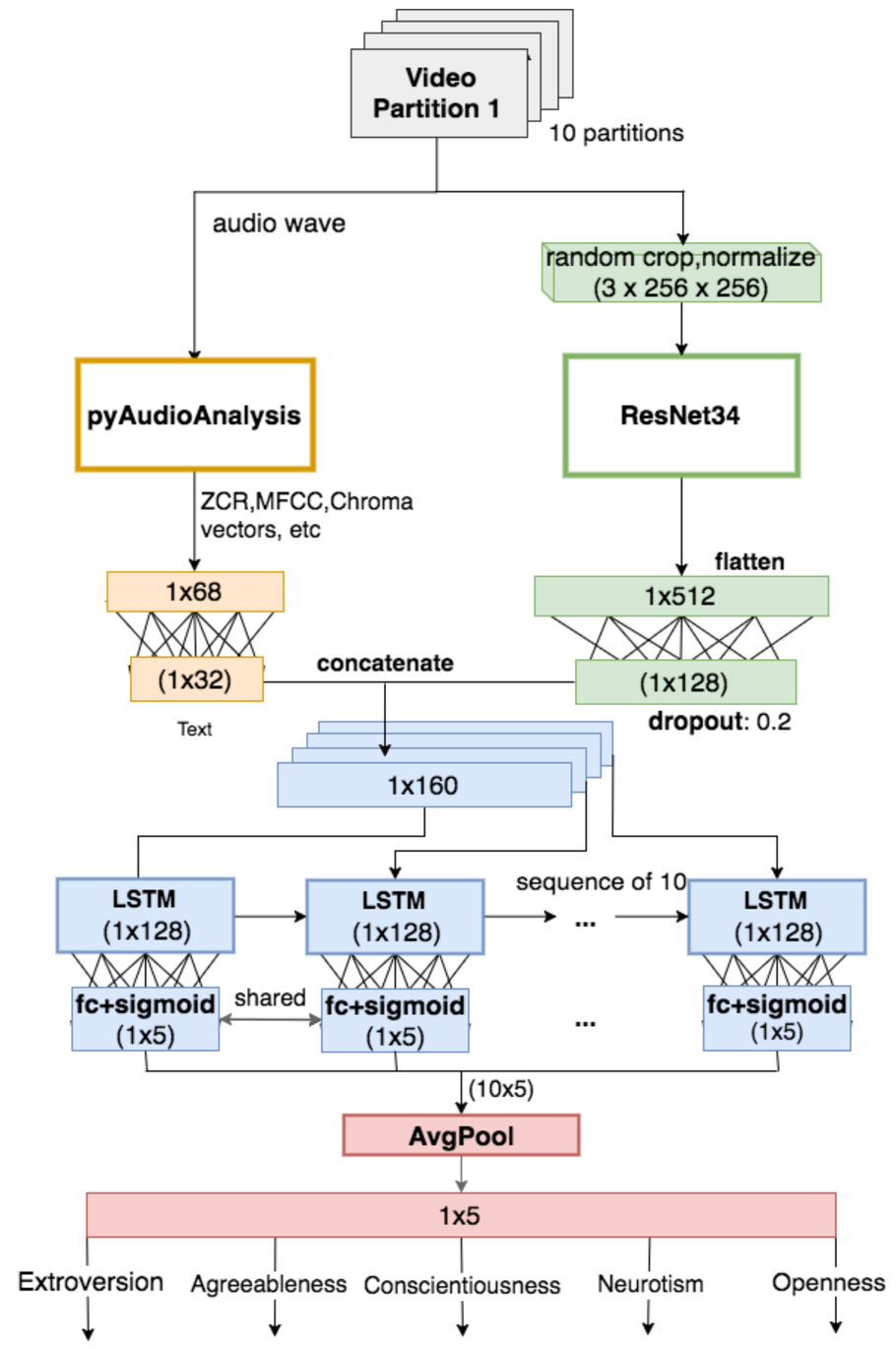

Figure 4. Architecture of Deep Bimodal Regression LSTM [1,2,3]

4. Results and Evaluation

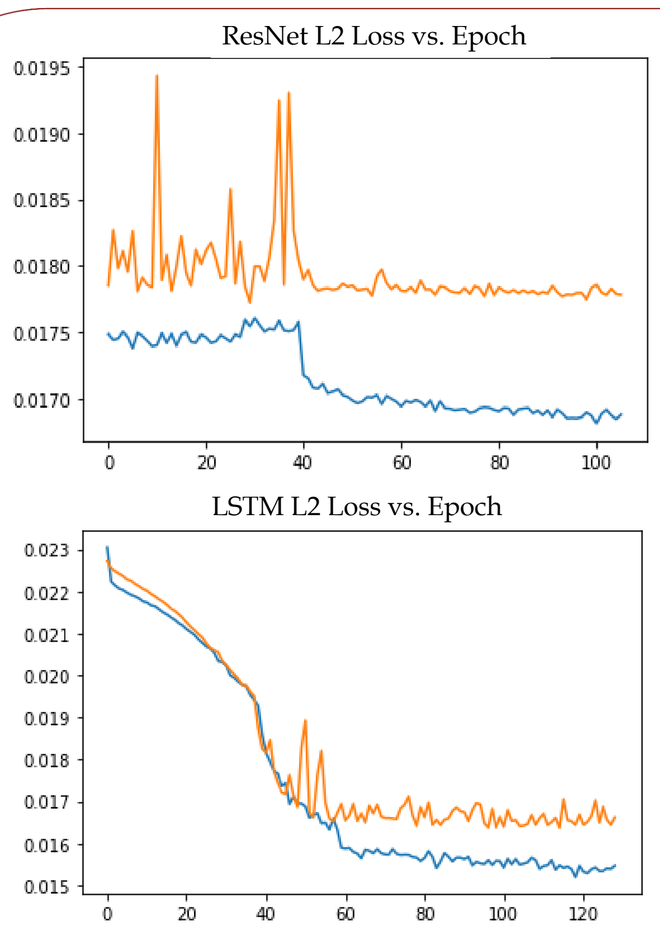

Figure 5. Training Process for ResNet and Bimodal LSTM

Evaluation Result						
Team	Mean Acc.	Extra.	Agree.	Concs.	Neuro.	Open.
Ours	0.9083	0.9110	0.8944	**0.9220**	0.9005	**0.9136**
NJU-LAMBDA	0.9130	0.9133	0.9126	0.9166	0.9100	0.9123
evolgen	0.9121	0.9150	0.9119	0.9119	0.9099	0.9117
DCC	0.9100	0.9107	0.9102	0.9138	0.9089	0.9111
ucas	0.9098	0.9129	0.9091	0.9107	0.9064	0.9099
BU_NKU	0.9094	0.9161	0.9070	0.9133	0.9021	0.9084

Table.1 Comparison of our results with the top 5 teams in ChaLearn First Impression Challenge [4,5,6]

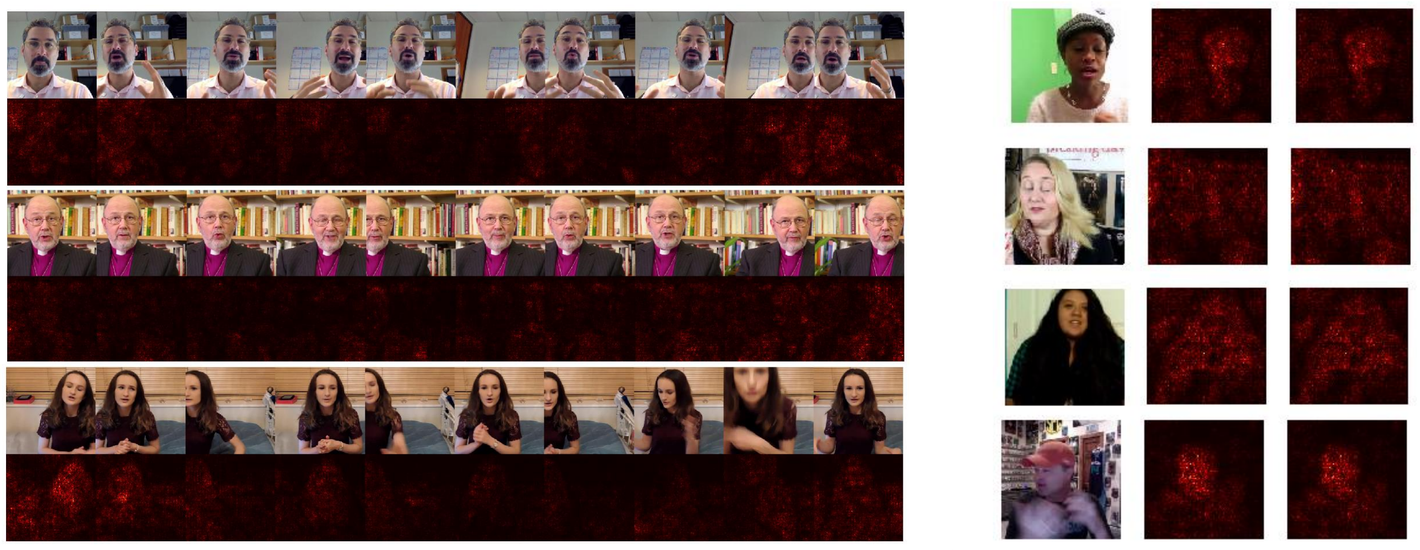

Figure 6. Extroversion score pixel saliency for Bimodal LSTM [left] and ResNet [right, col 1]. Both appear mostly influenced by human subjects. ResNet col 2 is saliency for neuroticism to show the trend is similar across features

5. Future Work

(a) Experiment with other visual feature CNN extractors (VGG, Inception, etc)
(b) Experiment with different RNN structures and input granularity in time
(c) Add a speech cue: use transcripts extracted by speech recognition.

6. Acknowledgements

We would like to thank the CS231N instructors Fei-Fei Li, Justin Johnson and Serena Young , as well as the TAs for their guidance and support.

References:

[1] Torchvision. Resnet Pytorch Pretrained Model. https://github.com/pytorch/vision/blob/ master/torchvision/models/resnet.py, 2017.
[2] T. Giannakopoulos. pyAudioAnalysis. https://github.com/tyiannak/pyAudioAnalysis, 2017
[3] P. Adam, S. Gross, S. Chintala. pyTorch. https://github.com/pytorch/pytorch, 2017
[4] A. Subramaniam, V. Patel, A. et al. Bi-modal first impressions recognition using temporally ordered deep audio and stochastic visual features. 2016
[5] C. Zhang, H. Zhang, X. Wei, et al. Deep bimodal regression for apparent personality analysis. 2016
[6] Y Gucluturk, U Guclu, M. van Gerven, R van Lier. Deep Impression: Audiovisual Deep Residual Networks for Multimodal Apparent Personality Trait Recognition. 2016
[7] J. Willis and A. Todorov. First impressions: Making up your mind after a 100-ms exposure to a face. Psychological Science, (17):592–598, 2006
[8] J.-I. Biel and D. Gatica-Perez. The youtube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. Multimedia, IEEE Transactions on, 15(1):41– 55, 2013. http://chalearnlap.cvc.uab.es/dataset/24/description/#.