Multimodal Prediction and Analysis of Latent User Dimensions

Laura Wendlandt and Rada Mihalcea
University of Michigan
{wenlaura,mihalcea}@umich.edu

Ryan L. Boyd and James W. Pennebaker
The University of Texas at Austin
{ryanboyd,pennebaker}@utexas.edu

Introduction

Research Questions:
- From a **correlational perspective**, how do image and caption attributes relate to personality and gender?
- Do image and caption attributes have **predictive power** for these traits?
- How can we use a **multimodal approach** to achieve better results?

Big 5 personality traits [2]:

Openness: artistic, curious, imaginative, insightful, original, wide interests

Conscientiousness: efficient, organized, planful, reliable, responsible, thorough

Extraversion: active, assertive, energetic, enthusiastic, outgoing, talkative

Agreeableness: appreciative, forgiving, generous, kind, sympathetic, trusting

Neuroticism: anxious, self-pitying, tense, touchy, unstable, worrying

Dataset and Features

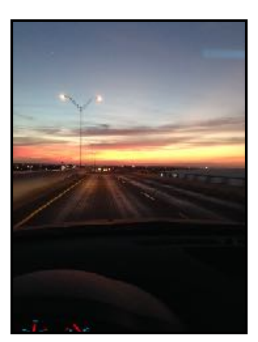
- Collected by Sam Gosling and James Pennebaker (UT Austin) from a Fall 2015 introductory undergraduate psychology class
- Includes five images, associated captions, gender, and personality
- Total: 1,353 students
- Sample images and captions:

"The real me is right behind you."

"Gotta find something to do when I have nothing to say."

"I'd rather be on the water."

"I crossed this bridge almost every day for 18 years and never got tired of it."



"The littlest things are always so pretty (and harder to capture)."

- **Image Features Extracted**
  - *Raw visual features* - colors, brightness and saturation, texture, static and dynamic lines, circles
  - *Scenes*
  - *Faces*
  - *Objects*

- **Caption Features Extracted**
  - *Stylistic features* - number of words, number of long words, named entities, readability, specificity
  - *N-grams*
  - *Part-of-speech n-grams*
  - *LIWC* (psychologically-based features) [4]
  - *MRC* (word statistics)
  - *word2vec embeddings* [3]

Methods

- Confusion matrix for text-predicted attributes and image-predicted attributes shows that images and text capture different aspects of personality

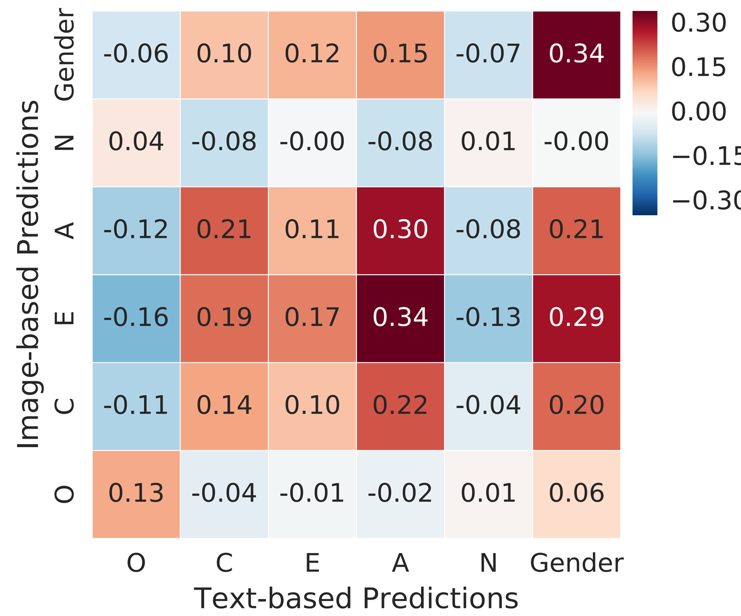


Image-Enhanced Unigrams (IEUs)
- Bag-of-words representation of **both** an image and its corresponding caption
- Includes all caption unigrams, as well as unigrams derived from the image
  - Any objects detected
  - Scene with the highest probability
  - Any color covering more than 33% of the image

Color: yellow, orange *Scene: kitchen*

Objects: plate, fork, salad, table

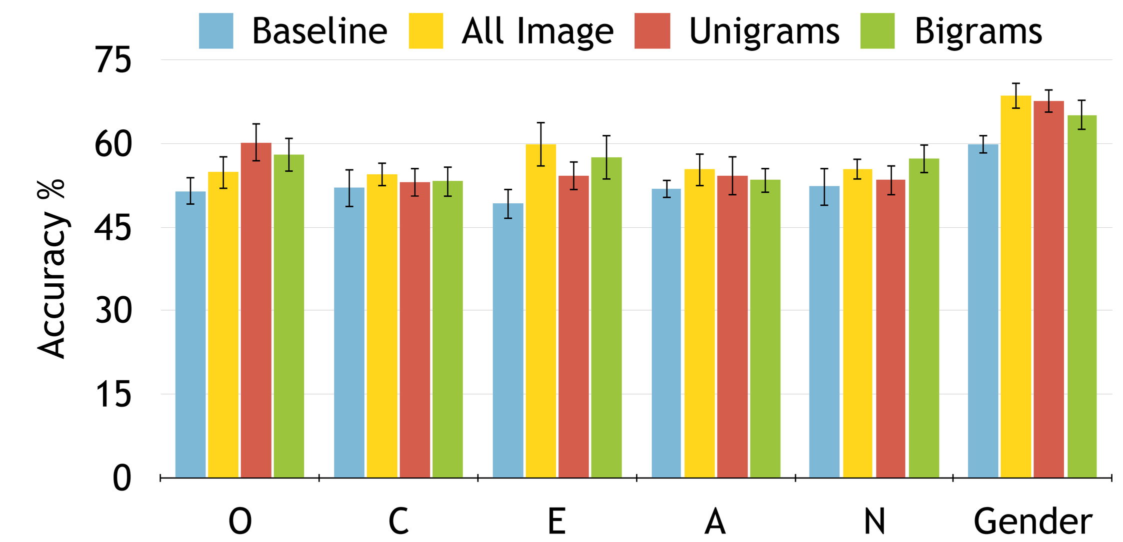
Macro v. Micro IEUs
- Macro IEUs:
  1. Extract unigrams from individual images
  2. Combine unigrams
- Micro IEUs:
  1. Extract and combine image attributes
  2. Extract unigrams from **combined** vector

Results

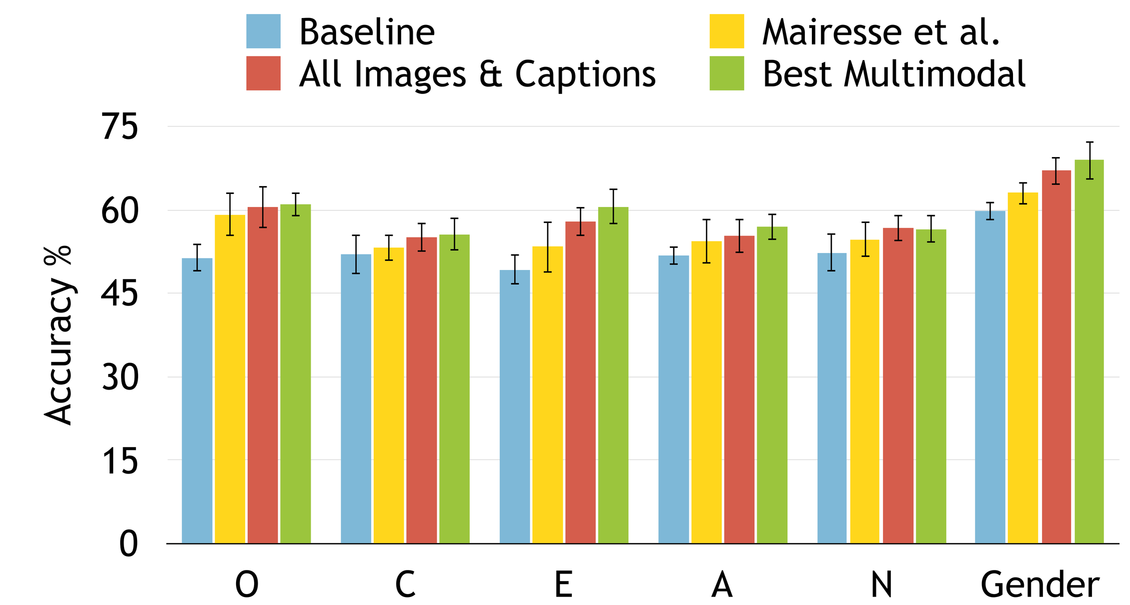
Classification Task
- Data divided into high segment and low segment for each trait (split at one standard deviation above/below mean)
- 10-fold cross validation on random forest with 500 trees
- Baseline: most common training class

Best Multimodal Method
- Average together pre-trained word2vec embeddings for all caption unigrams and all macro IEUs

Visual and Textual Features Only



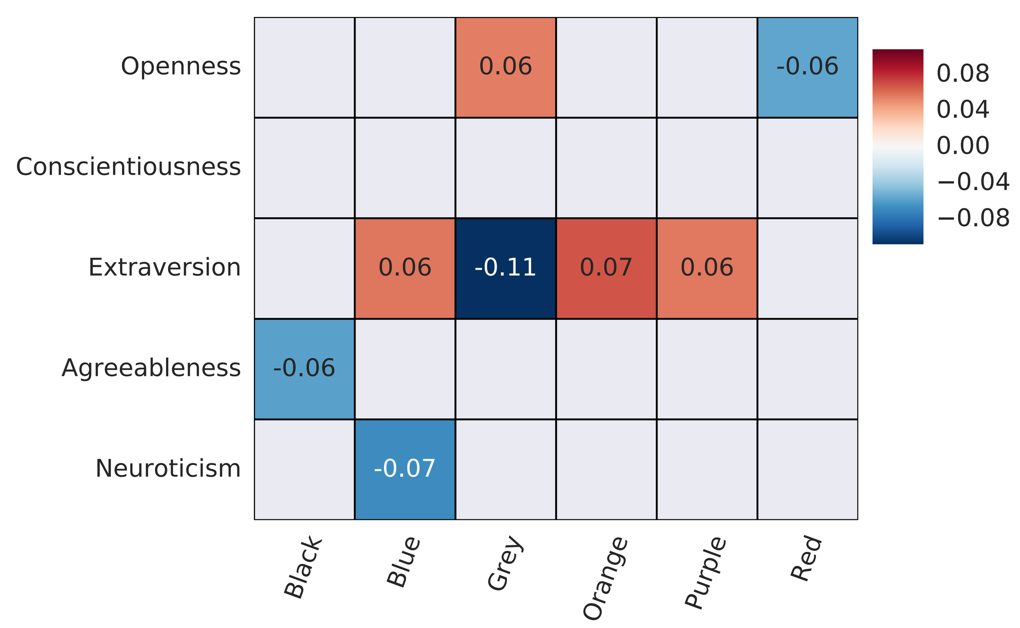
Multimodal

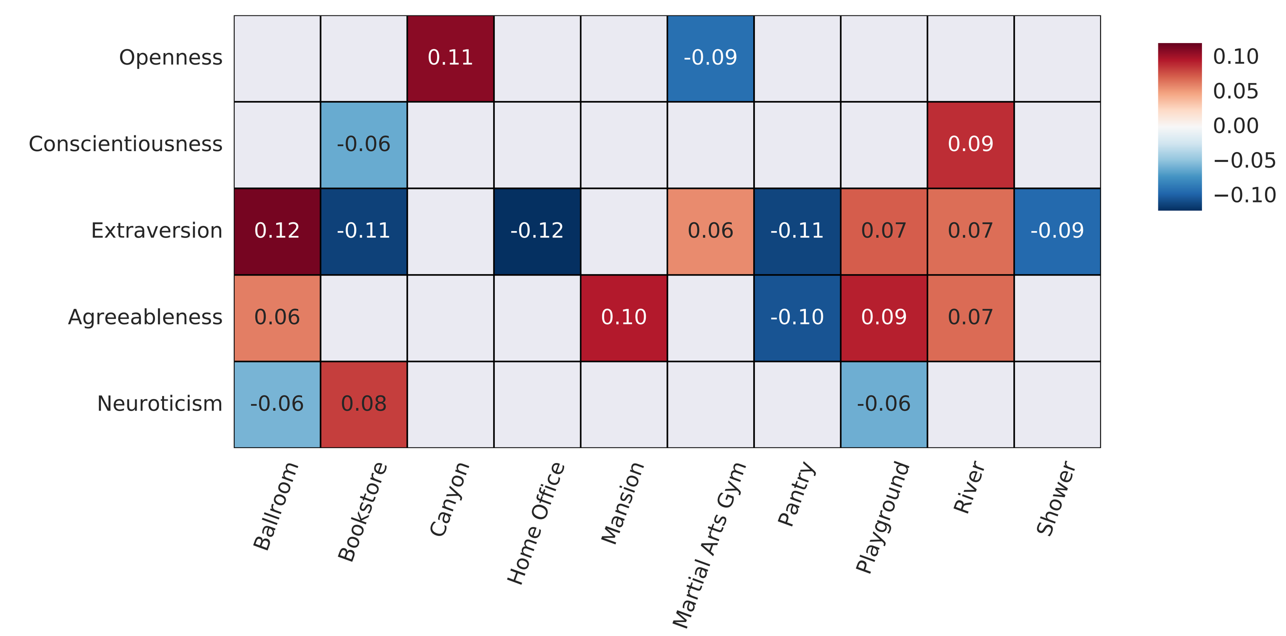


- Best multimodal method able to significantly predict openness, extraversion, agreeableness, and gender

Correlations

- Correlations calculated using a multivariate permutation test
- Pearson's r reported



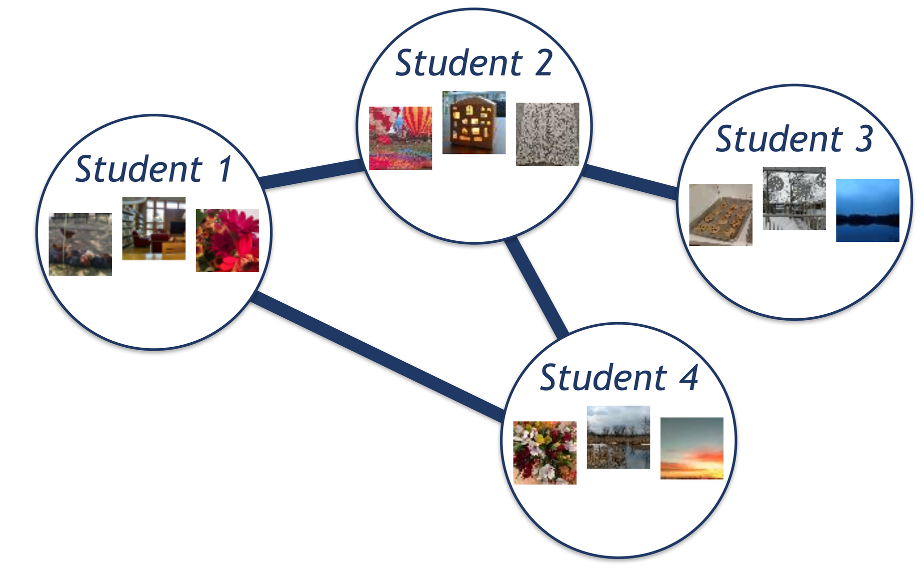


Conclusions

- Correlational techniques provide **interpretable psychological insight** into personality and gender.
- **Visual features** alone have significant predictive power.
- **Multimodal models** outperform both visual features and textual features in isolation, using a relatively small dataset.

Future Work

- Leverage **inherent network structure** in data to improve prediction



Outstanding Questions
- What is the best way to build a network from the dataset?
- What kind of network features will enhance prediction?
- How do you combine network features with image and text features?

References

[1] Mairesse, F.; Walker, M. A.; Mehl, M. R.; and Moore, R. K. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research* 30:457-500.

[2] McCrae, R. R., and John, O. P. 1992. An introduction to the five-factor model and its applications. *Journal of Personality* 60(2): 175-215.

[3] Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, 3111-3119.

[4] Pennebaker, J. W., and King, L. A. 1999. Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology* 77(6):1296.

Acknowledgements

This material is based in part upon work supported by the National Science Foundation (#1344257), the John Templeton Foundation (#48503), and the Michigan Institute for Data Science. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation, the John Templeton Foundation, or the Michigan Institute for Data Science.

We would like to thank Samuel Gosling for helping with the dataset collection, Shibamouli Lahiri for providing the code to calculate readability features, and Steven R. Wilson for providing the code to implement the Mairesse et al. paper that we use for prediction comparison.