

Yuhao Du

yuhaodu@buffalo.edu

716-207-0658

<https://yuhaodu.github.io/>

4445 Chestnut Ridge Road, Amherst, NY

Research Interest

Computational Social Science, Natural Language Processing, Machine Learning, Media Analysis, Algorithmic Fairness

Education

University at Buffalo	Ph.D. in Computer Science and Engineering Dean's Fellowship Advised by Dr. Kenneth Joseph	Sept. 2016 – Current
Massachusetts Institute of Technology	Visiting Graduate Student At Summer Institute of Computational Social Science	Summer 2019
Xi'an Jiaotong University	B.S. degree in Information and Computing Science	Sept. 2012 - June 2016

Technical Skills

- **Skills:** Machine learning; Deep learning; Natural Language Processing; Computer Vision; Probabilistic Graphical models
- **Packages:** PyTorch; TensorFlow; Keras; spaCy; NLTK; OpenCV; scikit-learn

Research Projects

Gender Bias in Corporations Fall 2020 – Current

- Used **agent-based model** to simulate a simplified workplace where two things happen: employees engage in projects and employees are promoted through the ranks of the company; Implemented influence of gender on perceived promotability of each employee.
- Found that in our simplified workspace, a 2 % bias can pile up and have drastic influence on gender representations.

Comparison between Linguistic Feature of Bodybuilding and Pro-Anorexic Content on Twitter Fall 2019 – Spring 2020

- Given the set of search terms labelled by domain expert, collected tweets related to bodybuilding, pro-ana and fitspiration.
- Built the classifier to distinguish tweets of different topics using **Bidirectional Encoder Representations from Transformers (BERT)**.
- Found that tweets related to fitspiration is more similar to tweets related to pro-ana than bodybuilding to pro-ana.

Bias in Word Embeddings Fall 2019 – Spring 2020

- Leveraged **Manifold Dimensionality Retention** to flatten the non-linear manifold structure in word embeddings; Identified the gender bias direction by incorporating information from the word vectors that are proximal to gendered pronouns; Removed the bias direction from word embeddings.
- Bias in upstream bias test is decreased from **0.925 to 0.53**. Bias in downstream bias test is decreased from **14.42 to 13.6**.

Understanding Visual Memes on Twitter Fall 2018 – Fall 2019

- Carried out the first large-scale study on the themes contained in the text of image-with-text (IWT) memes.
- Applied **Residual Neural network** to extract the visual features of user shared images.
- Took the average of **GloVe** word embedding for the text extracted by **Optical Character Recognition** from images as the textual features of the images.
- Fed the combined image and textual feature vectors into a **multi-layer perceptron** to classify IWT memes; The accuracy of the classifier is 78%.
- Found that 30% of the IWT memes in our sample which have identifiable theme are politically relevant.

Publications

- **Yuhao Du**, Muhammad Aamir and Kenneth Joseph. "Understanding visual memes: an empirical analysis of text superimposed on memes shared on Twitter." In *The Fourteenth International Conference on Web and Social Media (ICWSM)*, 2020.
- **Yuhao Du**, and Kenneth Joseph. "MDR Cluster-Debias: A Nonlinear Word Embedding Debiasing Pipeline." In *International Conference on Social Computing, Behavioral-Cultural Modeling, Prediction and Behavior Representation in Modeling and Simulation (SBP-BRIMS)*, 2020.

Professional Service

- **Conference Reviewer**
EMNLP 2019, ICWSM 2020, AAAI 2020
- **Conference Program Committee**
WebSci 2020