Machine Learning Classifiers for Socio-Demographics of Social Media Users
Challenges and Possibilities

Social media for public health makes possible massive, real-time analyses of health behaviors and opinions. However, **users don't report demographics,** requiring automated methods to contextualize these analyses. We explore **challenges in collecting data and training predictive models** for social media user demographics.

PREDICTIVE MODELS

1. Bag-of-words model of the user's writing history
2. Character-level neural network model of the user's (screen) name

LABEL COLLECTION

1. Buy (un)trained human labels
2. Scrape (noisy) online self-reports
3. Survey self-reports for evaluation

TAKEAWAYS

1. With enough data, can predict demographics from names alone
2. Helps to scrape noisy self-reports
3. All datasets raise concerns of bias

Label: human annotators look at entire Twitter profile

Label: "(he/him)" self-reports gender

Model: p(X | name="Zach")

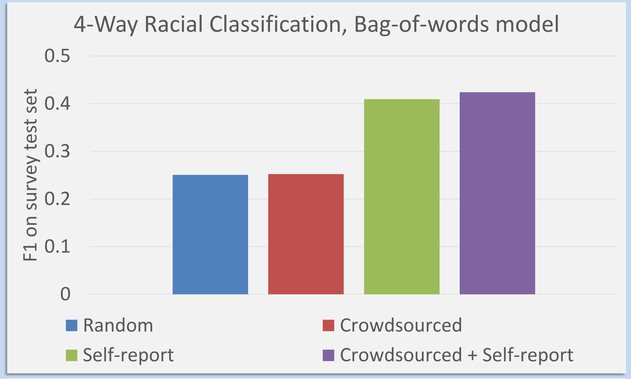
Model: p(X | words=["applying," "to," "grad," "school," …])

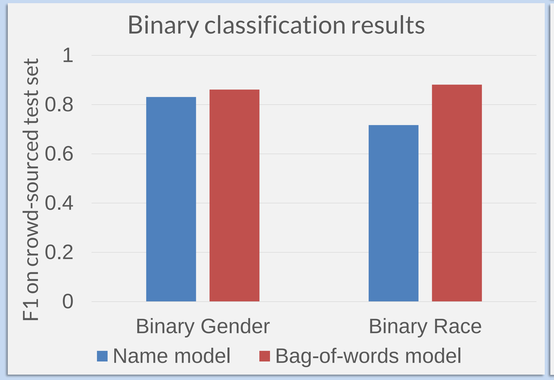
CHALLENGES

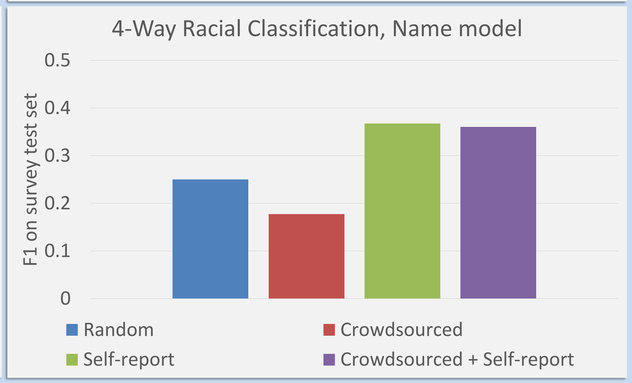
- How are demographic categories defined on social media? What are the ethical consequences of imperfect definitions?
- Do untrained annotators make systematic mistakes on different demographic groups?
- Are users who self-report demographics representative of the larger population?

POSSIBILITIES

- With enough data, our methods are accurate
- Can widen our data collection with automated methods
- Can explore or model possible selection bias







Zach Wood-Doughty, ♥ Mark Dredze, ♥
Michael Smith, ♦ David Broniatowski, ♦
Amelia Jamison, ♣ Sandra Quinn ♣
♥ Johns Hopkins University
♦ George Washington University
♣ University of Maryland

FUTURE WORK

1. Can we expand our data collection to better model the complexities of race and gender?
2. In what settings are these models accurate enough to contextualize large-scale public health analyses of social media data?
3. Can we mitigate methodological bias from human and automated labels?

REFERENCES

Wood-Doughty, Z., Smith, M., Broniatowski, D., & Dredze, M. (2017). How does twitter user behavior vary across demographic groups?. In *NLP+CSS* (pp. 83-89).

Wood-Doughty, Z., Andrews, N., Marvin, R., & Dredze, M. (2018). Predicting Twitter User Demographics from Names Alone. In *PEOPLES* (pp. 105-111).

Wood-Doughty, Z., Xu, P., Liu, X., & Dredze, M. (2019). Using Noisy Self-Reports to Predict Twitter User Demographics. *Under Review.*

This research was supported by the National Institute of General Medical Sciences, National Institutes of Health (NIH; award 5R01GM114771). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH.