Can digital footprints accurately predict political ideology? Evidence from Reddit

Michael Kitchener (32220162)
mkit0007@student.monash.edu

Supervisor: Assoc Prof. Paul Raschky
Paul.Raschky@monash.edu

May 13, 2021

What and why

- **What:** Using an original data set of 91,000 Reddit users, we are aiming to develop a predictive model of their political ideologies from their digital footprints
- **Why:** To illustrate that our online behaviour may implicitly disclose private information
  - If we can predict ideology from digital behaviour that is not explicitly political then individuals with democratic sympathies in authoritarian regimes may inadvertently reveal their views and risk harm.
  - The possibility of accurately estimating ideology may enable and encourage online political micro-targeting strategies such as voter suppression

What we know

- Neural networks can detect sexuality from pictures of a persons face (Wang and Kosinski, 2018)
- Facebook likes can be used to train statistical models capable of predicting personality traits to a higher degree of accuracy than close friends and family (Youyou et al., 2015)
- Natural language processing techniques can be used to accurately predict the political leanings of Twitter users (Colleoni et al., 2014)
- "Digital traces from social media can be studied to assess and predict theoretically distant psychosocial characteristics with remarkable accuracy" (Settanni et al., 2018)

What we're contributing

- These results pertain mostly to psychological traits
- Predictive power over political traits is salient to privacy concerns
- Our work is based on Kosinski's seminal work (Kosinski et al., 2013)
- We provide additional insight into this comparatively neglected area using a new, original data set to develop our understanding of these risks:
  - More powerful data (Likes can't distinguish degrees of interest)
  - More indicative of true interests (may not engage in some behaviours on Facebook, Twitter)
  - More sophisticated response variable
  - More advanced predictive methods

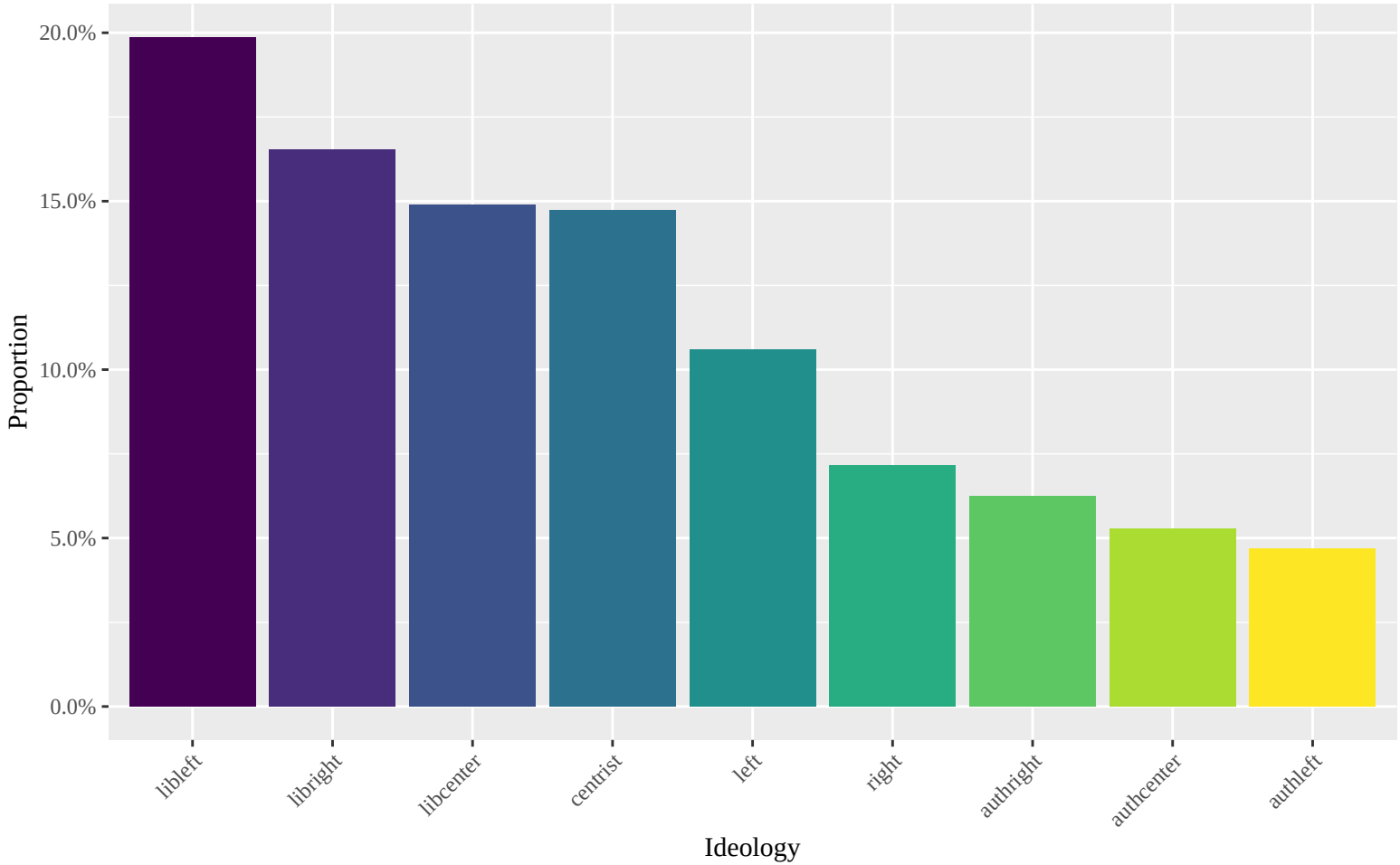
Methodology

- We collected usernames and flairs for 91,000 Reddit users who have flaired their results from the popular political compass test on the 'r/PoliticalCompassMemes' subreddit
- Users flair themselves with (typically) one of four ideologies
  - {left, authoritarian} i.e. communists
  - {right, authoritarian} i.e. traditional conservatives
  - {left, libertarian} i.e. social democrats
  - {right, libertarian} i.e. libertarians
- We are currently in the process of scraping the comment and post histories of these 91,000 users

User-flair data

username	**ideology**
user1	Libertarian-Left
user2	Authoritarian-Right
user3	Libertarian-Right
...	...

Ideology proportions (n=91,000)



User-history data

- Each post/comment saved as a row:
- Username | interaction | title | body | score | time | subreddit
- Easily transformed to:

username	**r/gaming**	**r/classicalmusic**	**r/boxing**	**r/seinfeld**
user1	3	12	0	0
user2	0	4	1	6
user3	0	0	0	0
user4	1	0	0	14
user5	0	43	0	0
...	...	...	...	...

Methodology (cont.)

- Merge this data with user-flair allowing us to model ideology as a function of digital footprint
- Remove columns referencing specifically political subreddits
- Data will have several 'irritating' features:
  - High dimensional: $p >> n$
  - Very sparse
  - Imbalanced
- As such, we will experiment with techniques for dimension reduction/variable selection (PCA, Lasso type penalties, etc.) and techniques for imbalanced data (under-sampling, optimizing for balanced accuracy, etc.)
- Specific models TBD as I need to look further into the best methods for classification with high dimensional, sparse data

Next steps

- Complete user-history scrape
- Train a variety of statistical models
- Examine results and interpret where possible
- Time permitting, illustrate the usefulness of Reddit data by examining changes in sentiment on politicised topics for out of sample predictions

Questions & comments!

- Thank you for listening!
- Any questions or comments?

References

Colleoni, E., Rozza, A., and Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of communication*, 64(2):317–332.

Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805.

Settanni, M., Azucar, D., and Marengo, D. (2018). Predicting individual characteristics from digital traces on social media: A meta-analysis. *Cyberpsychology, Behavior, and Social Networking*, 21(4):217–228.

Wang, Y. and Kosinski, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of personality and social psychology*, 114(2):246.

Youyou, W., Kosinski, M., and Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4):1036–1040.