gesis Leibniz Institute for the Social Sciences

Measuring Public Opinion Online

Potentials and Pitfalls

Indira Sen, Doctoral Candidate
08.06.2020

This talk

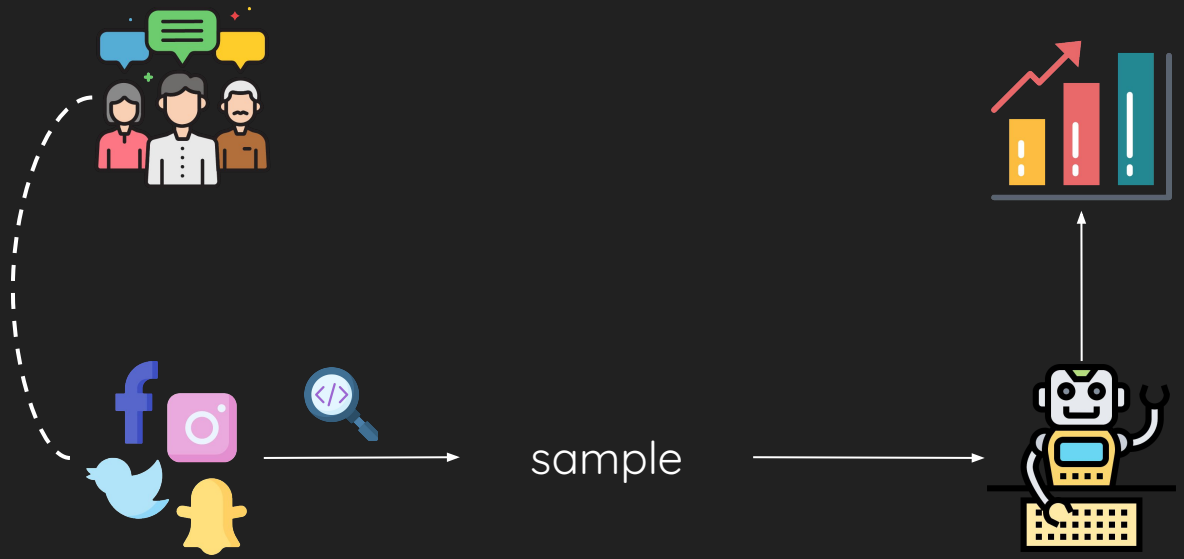
- How is public opinion measured online?

- What is the validity of ***theoretical proxies*** that can be used to measure public opinion online?

- What is the validity of the ***practical proxies*** that can be used to measure public opinion online?

This talk

- How is public opinion measured online?



How do we measure it?

O'Connor, B., Balasubramanyan, R., Routledge, B.R. and Smith, N.A., 2010, May. **From tweets to polls: Linking text sentiment to public opinion time series.** In *Fourth international AAAI conference on weblogs and social media.*

Pasek, J., McClain, C.A., Newport, F. and Marken, S., 2019. **Who's Tweeting About the President? What Big Survey Data Can Tell Us About Digital Traces?**. Social Science Computer Review, p.0894439318822007.

Conrad, F.G., Gagnon-Bartsch, J.A., Ferg, R.A., Schober, M.F., Pasek, J. and Hou, E., 2019. **Social media as an alternative to surveys of opinions about the economy.** Social Science Computer Review, p.0894439319875692.

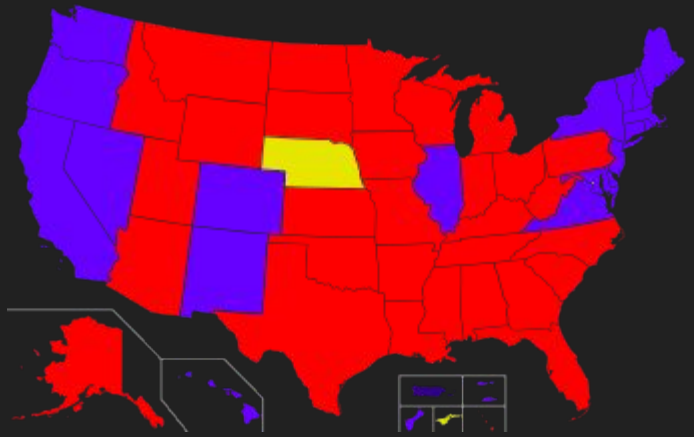
Findings from Previous Studies

- After some initial positive results, turns out social media measures =/= polls measures
- But why?

Findings from Previous Studies

- After some initial positive results, turns out social media measures =/= polls measures
- But why?
- Usually credited to **Representation errors**

Demographic breakdown of USA (political affiliation)



Errors?

- Representation: Errors due to ***who*** is measured
  - Coverage: demographic mismatch between social media and census
  - Non-response: certain people are not very vocal
  - Can be mitigated using **reweighting** [Barber á et al 2016]

Errors?

- Representation: Errors due to ***who*** is measured

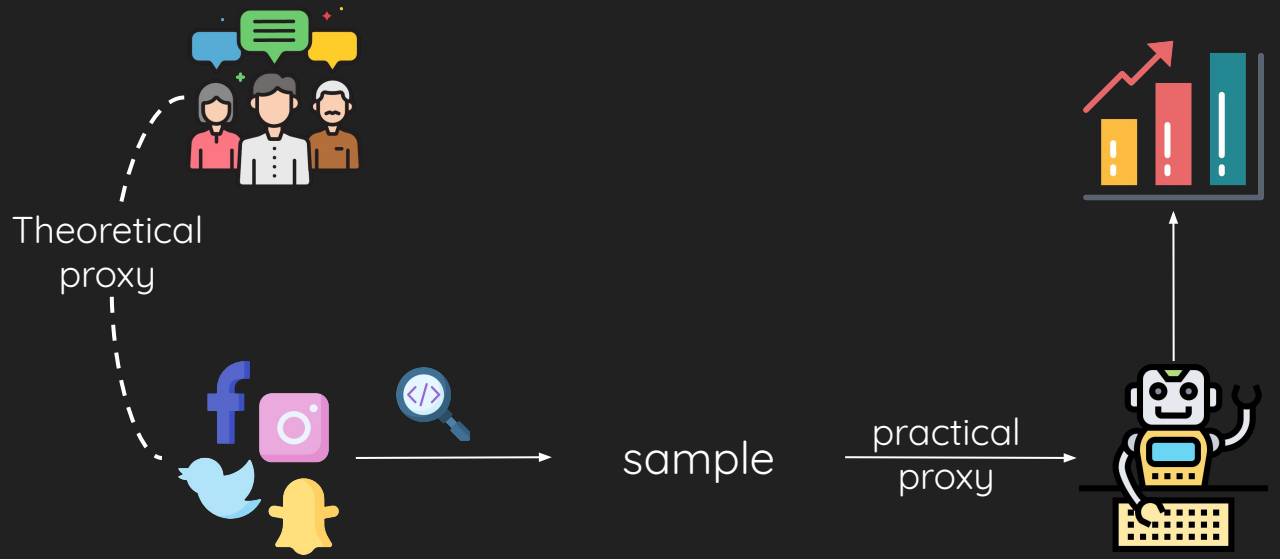
- But, what about **measurement** errors, i.e., errors due to ***how*** the construct is measured?

Sen, I., Floeck, F., Weller, K., Weiss, B. and Wagner, C., 2019. **A Total Error Framework for Digital Traces of Humans.** arXiv preprint arXiv:1907.08228

Diagnosing errors in computational social science studies

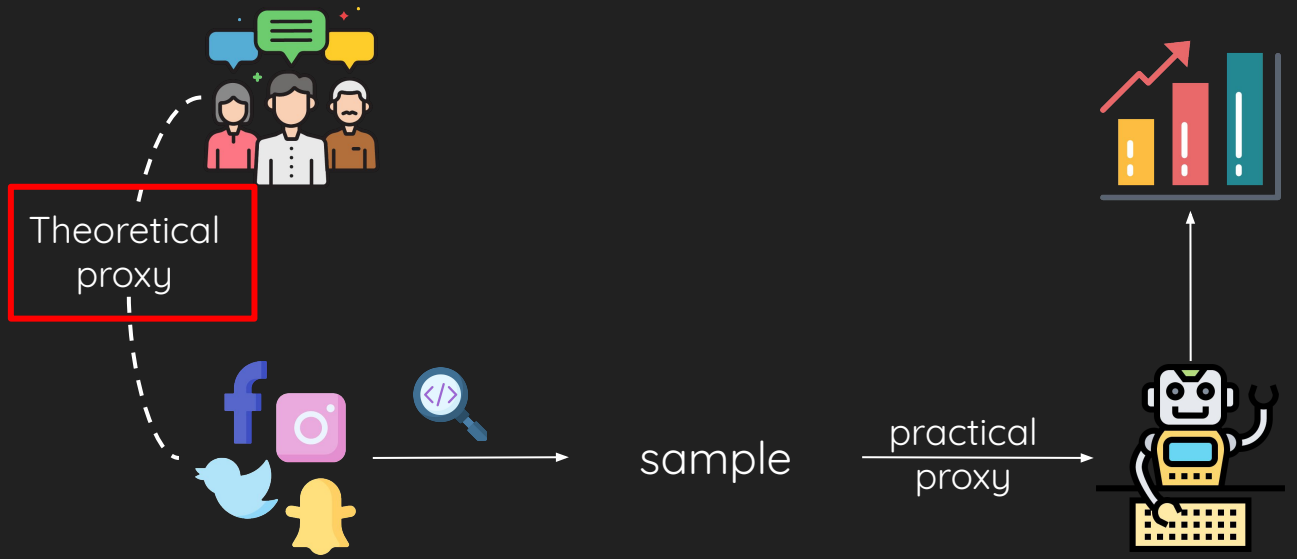
This talk

- How is PO measured online?



This talk

- How is PO measured online?



How do we measure ‘opinion’?

O'Connor, B., Balasubramanyan, R., Routledge, B.R. and Smith, N.A., 2010, May. **From tweets to polls: Linking text sentiment to public opinion time series.** In *Fourth international AAAI conference on weblogs and social media*.

Conrad, F.G., Gagnon-Bartsch, J.A., Ferg, R.A., Schober, M.F., Pasek, J. and Hou, E., 2019. **Social media as an alternative to surveys of opinions about the economy.** Social Science Computer Review, p.0894439319875692.

Sentiment

This talk

- How is public opinion measured online?

- What is the validity of ***theoretical proxies*** that can be used to measure public opinion online?

- What is the validity of the ***practical proxies*** that can be used to measure public opinion online?

Are we measuring it right? Validity

	POS	NEG	NEUT	Accuracy
opposing Brown	124	76	150	21.71%
opposing Coakley	70	67	105	27.68%
supporting Brown	216	45	254	41.94%
supporting Coakley	240	72	213	45.71%
neutral	249	82	296	47.20%
				36.85%

TABLE IV

CONFUSION MATRIX FOR THE EVALUATION OF THE AUTOMATIC SENTIMENT ANALYSIS COMPUTED AGAINST A MANUALLY LABELED SET OF TWEETS.

Metaxas, P.T., Mustafaraj, E. and Gayo-Avello, D., 2011, October. **How (not) to predict elections.** In 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing (pp. 165-171). IEEE.

Are we measuring it right? Validity

- Usually opinion is defined as ‘sentiment’

Tweet	Untargeted Sentiment	Opinion
Trump is the only candidate I fully support	positive	approval

Researcher wants measure opinion about **Donald Trump**

Are we measuring it right? Validity

- Usually opinion is defined as ‘sentiment’
- This sentiment is untargeted, therefore, if multiple entities are mentioned, sentiment =/= opinion

Tweet	Untargeted Sentiment	Targeted Sentiment	Opinion
What makes me angry is the media unnecessarily attacking Trump	negative	positive	approval

Researcher wants measure opinion about **Donald Trump**

Are we measuring it right? Validity

- Usually opinion is defined as ‘sentiment’
- This sentiment is untargeted, therefore, if multiple entities are mentioned, sentiment =/= opinion
- Finally, targeted sentiment doesn’t work when entities are ***not directly*** mentioned

Tweet	Untargeted Sentiment	Targeted Sentiment	Stance	Opinion
Jeb Bush is the only choice in the republican lineup	positive	none	against	disapproval

Researcher wants measure opinion about **Donald Trump**

What is ‘stance’?

- Rumor stance, argument stance, **viewpoint** stance

- **Indirect stance:** “the target is referred to in indirect ways such as through pronouns, epithets, honorifics, and relationships.” [Mohammad et al, 2017]

Stance as a theoretical proxy for Opinion

Tweet	Untargeted Sentiment	Targeted Sentiment	Stance	Opinion
Trump is the only candidate I fully support	positive	positive	favor	approval
What makes me angry is the media unnecessarily attacking Trump	negative	positive	favor	approval
Jeb Bush is the only choice in the republican lineup	positive	none	against	disapproval

This talk

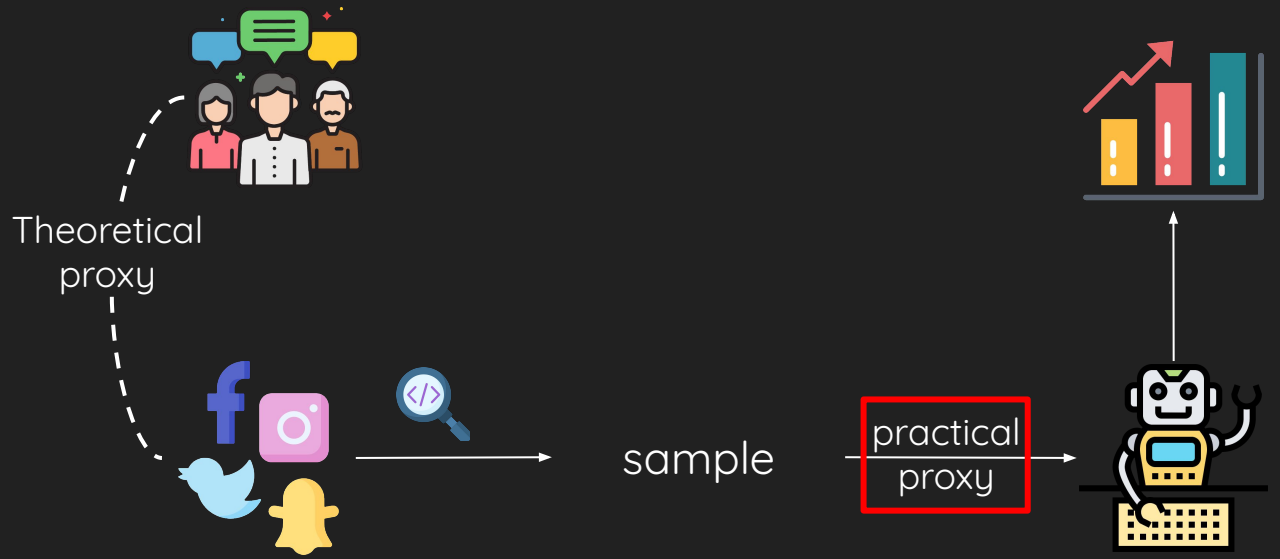
- How is public opinion measured online?

- What is the validity of ***theoretical proxies*** that can be used to measure public opinion online?

- What is the validity of the ***practical proxies*** that can be used to measure public opinion online?

This talk

- How is public opinion measured online?



Are we measuring it right *in practice*?

- How are stance methods in practice?

- Can they be used off-the-shelf like VADER can for sentiment?

Proposition: Benchmark methods

Proposition: Benchmark methods

- Train text classification models on ‘weak labels’
  - SemEval 2016 Stance Detection Shared Task: measuring opinion towards Donald Trump [Mohammad et al, 2016]
  - Strong method trained classifier on ‘weak labels’: using heuristics
  - Strong performance on SE test set

Proposition: Benchmark methods

- Train text classification models on ‘weak labels’: **BERT-SD**
- Test sets:
  - Experiment 1: multiple unseen targets
  - Experiment 2: one seen target

- Baselines:
  - Random classifier: randomly assign one from [‘favor’, ‘against’, ‘none’]
  - VADER

Datasets

Seen targets

Unseen targets

Target	Dataset	Tweets
Trump	SemEval, MTSD, PRES, Constance	3600
Clinton	SemEval	300
Macron	PRES	600
Putin		600
Zuma		600

Preliminary Results

Metric: Macro F1

Generalizability to Unseen Targets

- poor

Method	**Unseen targets**
Random	31.0
VADER	**45.3**
BERT-SD	40.2

Performance on a Seen Target?

- Familiar target: Trump
- Better

Method	Trump (SemEval)
Random	32.2
VADER	36.3
BERT-SD	**60.5**

Performance on a Seen Target?

- Familiar target: Trump
- Poor on other datasets

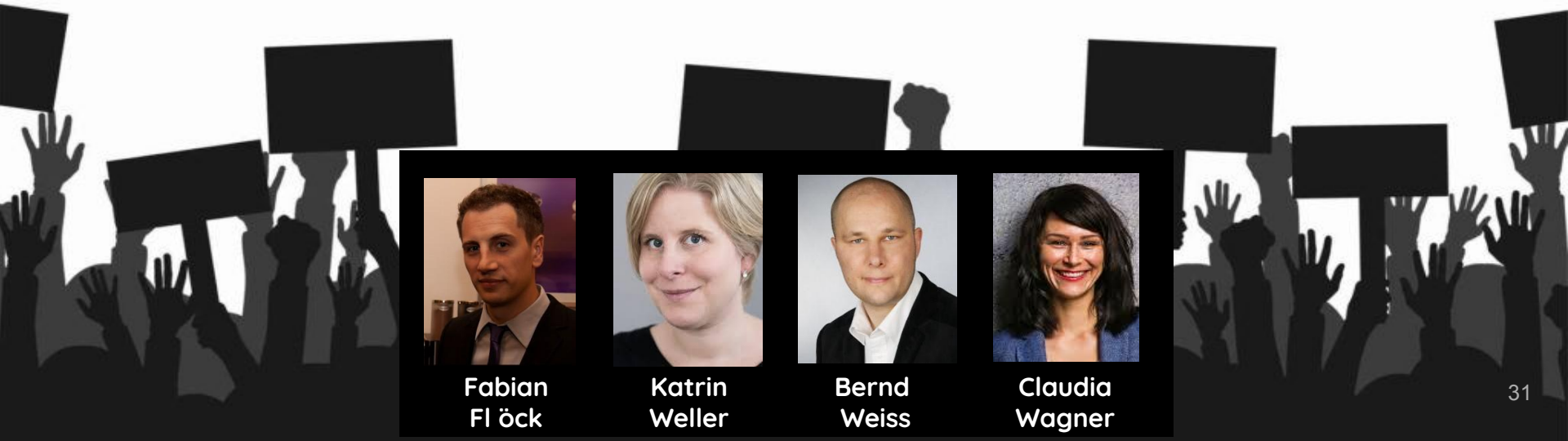
Method	Trump (SemEval)	Trump (Other)
Random	32.2	31.5
VADER	36.3	35.8
BERT-SD	**60.5**	42.6

Takeaways

- Need to look at **representation** errors, as well as **measurement** errors
- Preliminary results from benchmarking 14 sentiment and stance methods
  - Current methods have room for improvement
  - Need for generalizable methods

Diagnosing errors in CSS studies: A Total Error Framework for Digital Traces of Humans. arXiv:1907.08228.
https://tinyurl.com/ted2020

Email: **Indira.Sen@gesis.org**
Twitter: @indiiigosky



Fabian Fl öck

Katrin Weller

Bernd Weiss

Claudia Wagner

References

- Barberá, P., 2016. Less is more? How demographic sample weights can improve public opinion estimates based on Twitter data. *Work Pap NYU*.
- Mohammad, S.M., Sobhani, P. and Kiritchenko, S., 2017. Stance and sentiment in tweets. ACM Transactions on Internet Technology (TOIT), 17(3), pp.1-23.
- Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X. and Cherry, C., 2016, June. Semeval-2016 task 6: Detecting stance in tweets. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016) (pp. 31-41).
- [MTSD] Sobhani, P., Inkpen, D. and Zhu, X., 2017, April. A dataset for multi-target stance detection. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers (pp. 551-557).
- [Constance] Joseph, K., Friedland, L., Hobbs, W., Tsur, O. and Lazer, D., 2017. Constance: Modeling annotation contexts to improve stance classification. arXiv preprint arXiv:1708.06309.
- [PRES] van den Berg, E., Korfhage, K., Wiegand, M. and Ruppenhofer, J., 2019. Not my president: how names and titles frame political figures.