Bias and Methods of AI technology studying Political Science

Moa Johanson, Chalmers University of Technology

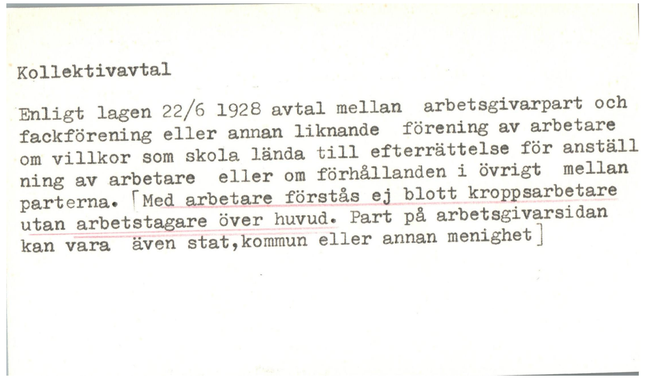
Joint work with:
Annika Fredén, Pasko Kisic Merino (KAU) and Denitsa Saynova (Chalmers)

CLASP Seminar 27 May 2021

AI in political science

- **Political Science questions:**
  - Coalitions forming/breaking up.
  - Upcoming policy changes.
  - Emerging political issues.
  - Voter behaviour.
- **“Text-as-data”**:
  - Twitter posts
  - Transcripts of speeches
  - Written parliamentary motions
  - Surveys

- Data source and choice of method(s)?
- How does choice of ML method affect results?
- Differences if we study English language countries vs. smaller languages?



Kollektivavtal

Enligt lagen 22/6 1928 avtal mellan arbetsgivarpart och fackförening eller annan liknande förening av arbetare om villkor som skola lända till efterrättelse för anställning av arbetare eller om förhållanden i övrigt mellan parterna. [Med arbetare förstås ej blott kroppsarbetare utan arbetstagare över huvud. Part på arbetsgivarsidan kan vara även stat,kommun eller annan menighet]

Word embeddings from different parties

Exploratory pilot study presented at PolMeth Europe March 2021

- Learn word meaning as numeric vector - a word embedding.
- Similar meaning/use to give similar embeddings.
- How does embeddings differ when **trained on data from different parties**?




Some Related Work

- **Rodman** (2020): *A Timely Intervention: Tracking the Changing Meanings of Political Concepts with Word Vectors*
  - word2vec embeddings to study meaning over time, ideological bias.
  - E.g. “Equality” in news article headlines 1855-2016.
  - Comparison to topic modelling - broad themes vs word meaning.
- **Rheault & Cochrane**: *Embeddings for Political Analysis*
  - Parliamentary speeches (US, Canada, UK).
  - “Party embeddings” for each party and session of parliament.

Sweden and the US

- Sweden:
  - Proportional representation, coalitions.
  - Left - Right
  - Written motions from parliament.
- US:
  - Presidential and individual candidate elections.
  - Liberal - Conservative
  - Transcripts of Senate speeches.

Data

- **US:**
  - Parliamentary debates (2015-2017)
  - Republicans: 19K docs, 8.5M tokens
  - Democrats: 17K docs, 8.4M tokens
- **Sweden:**
  - Written motions (2013-2017)
  - Moderates: 6K docs, 3M tokens
  - Social Democrats: 4K docs, 2M tokens.

Methods

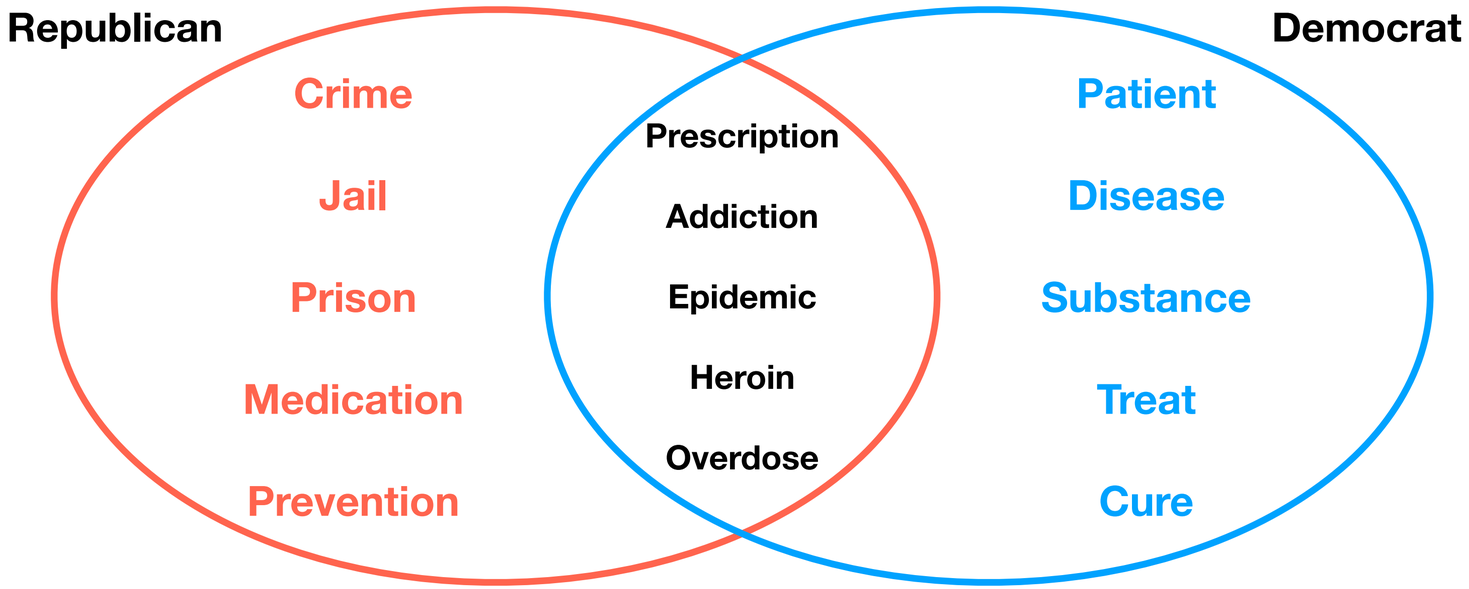
- Count-based:
  - PMI: co-occurrence matrix normalised by number of occurrences.
  - SVD: compressed PMI matrix (reduce noise)
- Neural Networks (word2vec):
  - CBOW: Learn embedding by training to predict missing word in given context.
  - Pre-training + fine tuning

Topics: Taxes and Drugs

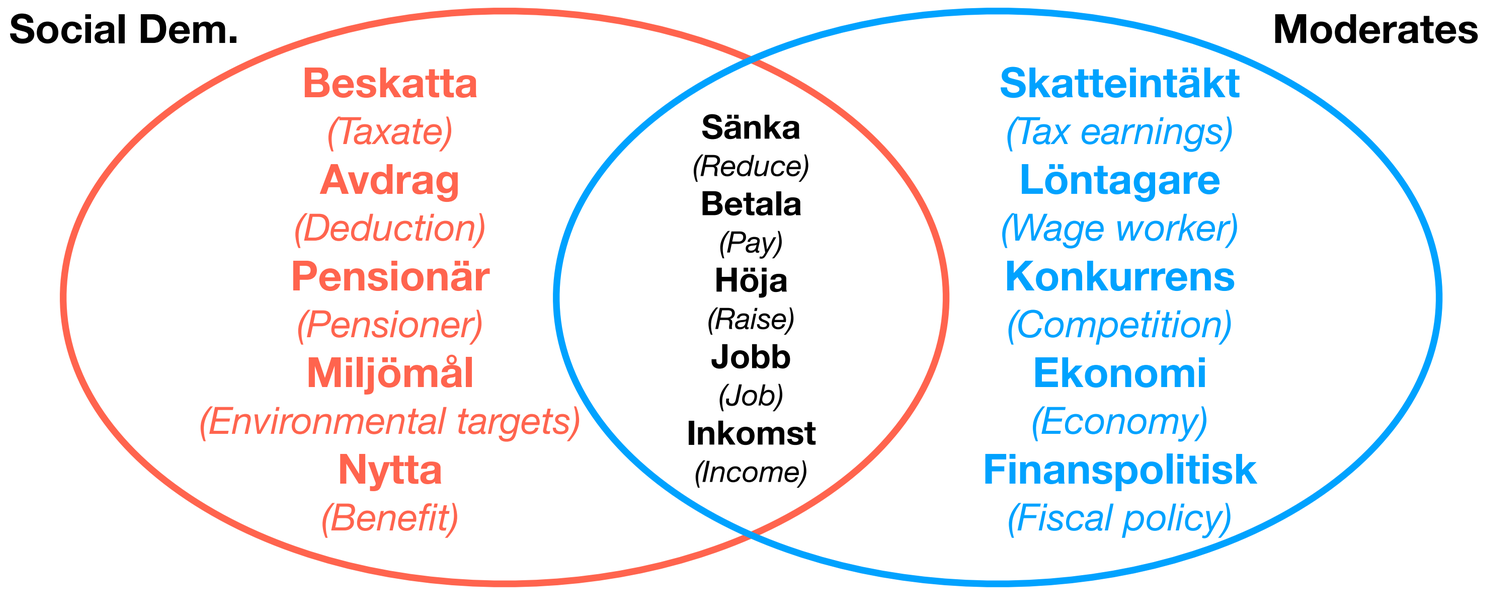
- **Taxes:**
  - Expect division along left-right dimension.
- **Drugs:**
  - Not clearly a left-right issue.

- Measure cosine distance to these seed-words - find closest. Do they differ depending on which party used for training?

Sample results: Drugs (US)



Sample result: Taxes (Sweden)



Summary

- Drugs topic similar pattern in US and Sweden.
- Taxes dimension more pronounced in Sweden.
- PMI model gave more interesting results detecting nuances in agreement/ disagreement in word meaning.
- Neural Network tended to give example words (eg. listing different kinds of drugs).
    - Due to small window size, while PMI summarise whole document.

Next Steps

Data Science/ML: Explainable ML

- Political Science: mainly un-supervised problems.
  - How do we know our ML models are any good?
- Re-cast as **supervised learning**: learn which party wrote motion.
  - Test also large NN models, transformers, BERT.
- **Explainability techniques:** NN as black box, explain what characterise each party?
- Focus on **Swedish parliamentary motions**.
- Dealing with classifications of long documents - hierarchical NN's?

Next Steps

Political Science

- Analysis of “political influencers” in social media.
- First step: Brexit - qualitative analysis of posts by influencers, populists, parliamentarians.
- Social network analysis, polarisation, filter bubbles.
- Could serve as gold-standard for later quantitative analysis.

Thanks for listening!

Psst!
Post-doc position will be announced soon!