From Talk to Action with Accountability:

Monitoring the Public Discussion of Policy Makers with Deep Neural Networks and Topic Modelling

ICML 2021

Tackling Climate Change with Machine Learning
Proposals Track

People Behind the Proposal

Vili Hätönen

Data scientist at

Emblica

vili@emblica.com

Fiona Melzer

Cognitive Science at

University of Edinburgh

F.Melzer@sms.ed.ac.uk

Mentor

Malte Hessenius

The Main Premise

We already have the tools to mitigate climate change.

The problem is that **we are not using those tools**.

Table of Contents

ML relevance

CC relevance

1. Accountability & MuSTAS

The Problem

- Decision makers talk about actions to mitigate climate change
- Mismatch between promises & actions taken
- To increase actions we need **accountability**

Accountability is Crucial

Increased accountability of government officials has been shown to be crucial in preventing political mismanagement.

The degree of information citizens have about their government's actions is one of two major hinges in accountability (Adsera et al., 2003).

⇒

Good **degree of information** is a requirement for accountability

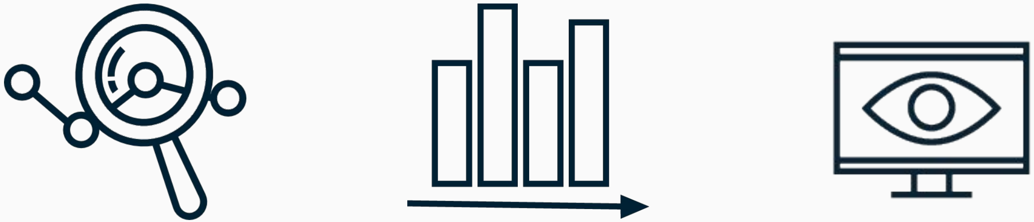
Access to Data is not Enough

- Today there is more publicly available data on policy makers than ever before.

- Voters, NGOs, or watchdog organisations can't be expected to analyse
    - **all** data, from
    - **all** the sources
    - **all** the time.

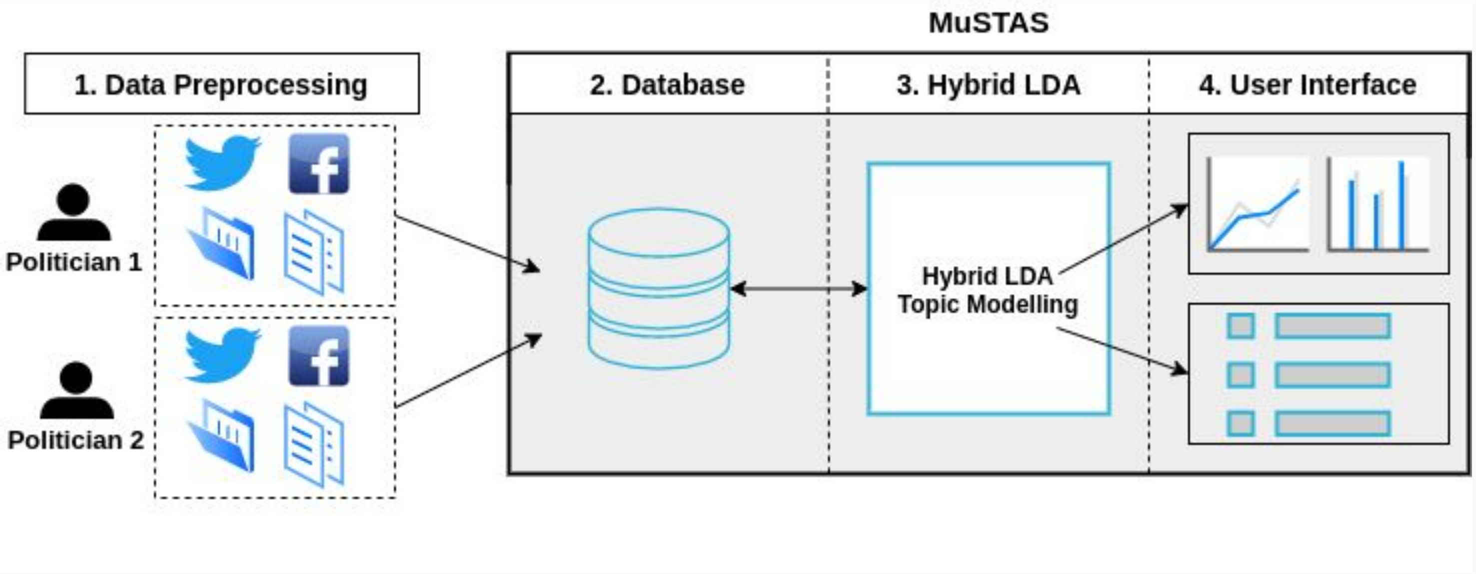
- Outsourcing information gathering to major news outlets does not solve the problem.

Solution



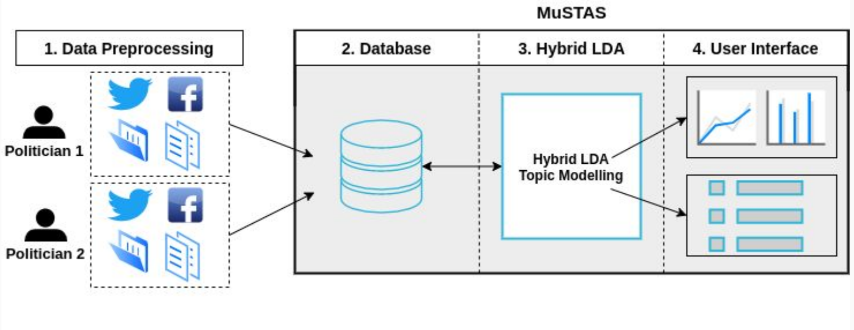
Provide voters & civil society with a tool of insight into policy makers' talk

Multi-Source Topic Aggregation System (MuSTAS)



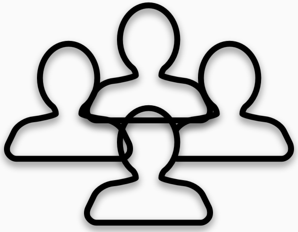
MuSTAS Benefits

- Process policy makers' speech from
  - several sources, also the ones you wouldn't normally follow
  - daily,
  - for e.g. all the members of parliament for all the parties.
- Distill the information to concise **topics**
  - an easily digestible format that provides an overview
- Provide the digest in an easily accessible format for all the stakeholders to use
  - with links to the original source documents



Identified User Groups

Voters



Civil Society

MuSTAS for Voters

Voters

Problems:

- Traditional media has biased coverage per
  - party
  - candidates
  - timeline
  - topic

- Voters get politically divided depending on source of coverage

- Politician’s narrative differs between media

MuSTAS for Voters

Voters

MuSTAS solution:

- get analysis per
  - party
  - candidate
  - timeline
  - topic
  - medium
- access to narrative on media the voters don’t follow

➔ *Does my candidate discuss in parliament what they claim on media?*

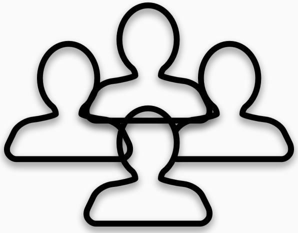
➔ *Does my candidate / party discuss different topics on different media?*

➔ *How does my candidates topics differ before and after being elected?*

➔ *How do different parties differ in climate topics brought up in parliamentary discussions?*

➔ *What are the parties’ political narratives on platforms I don’t follow?*

MuSTAS for Civil Society

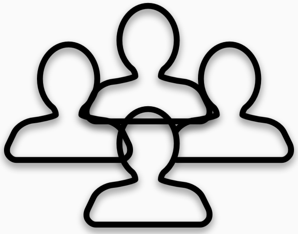


Problems:

- large scale analysis of political talk is costly
- previous analysis only tackles one medium at a time
  - legislative speech US Senate 1997-2004 [1]
  - public opinion survey response UK 2021 [2]
  - public policy analysis India 2020 [3]
- quantity of source data is overwhelming

[1] Quinn, Kevin & Monroe, Burt & Colaresi, Michael & Crespin, Michael & Radev, Dragomir. (2010). How to Analyze Political Text With Minimal Assumptions and Costs. American Journal of Political Science. 54. 209 - 228.
[2] Wright, L. *et al.* (2021) 'Public Opinion about the UK Government during COVID-19 and Implications for Public Health: A Topic Modelling Analysis of Open-Ended Survey Response Data', *medRxiv*, p. 2021.03.24.21254094. doi: 10.1101/2021.03.24.21254094.
[3] Debnath R, Bardhan R (2020) India nudges to contain COVID-19 pandemic: A reactive public policy analysis using machine-learning based topic modelling. PLOS ONE 15(9): e0238972.

MuSTAS for Civil Society



Civil Society

MuSTAS solution:

- Integrated analysis of political discourse across platforms
  - twitter, facebook, personal blogs
  - official interviews, parliamentary discussions

- Facilitated research by providing source-documents
  - tagged by source, candidate, topic, medium

- Easy adoption for other parliamentary democracies
  - needs model for target-language
  - needs source text , social media handles

2. Multi-Source Hybrid LDA

Requirements for Topic Model used in MuSTAS

- Reliably model topics over documents with **different styles**:
  - different vocabulary
  - different length documents
  - topics are not evenly distributed between different document types

- Reliably model topics with **context shift**
  - Model trained with today's data should work in 6 months

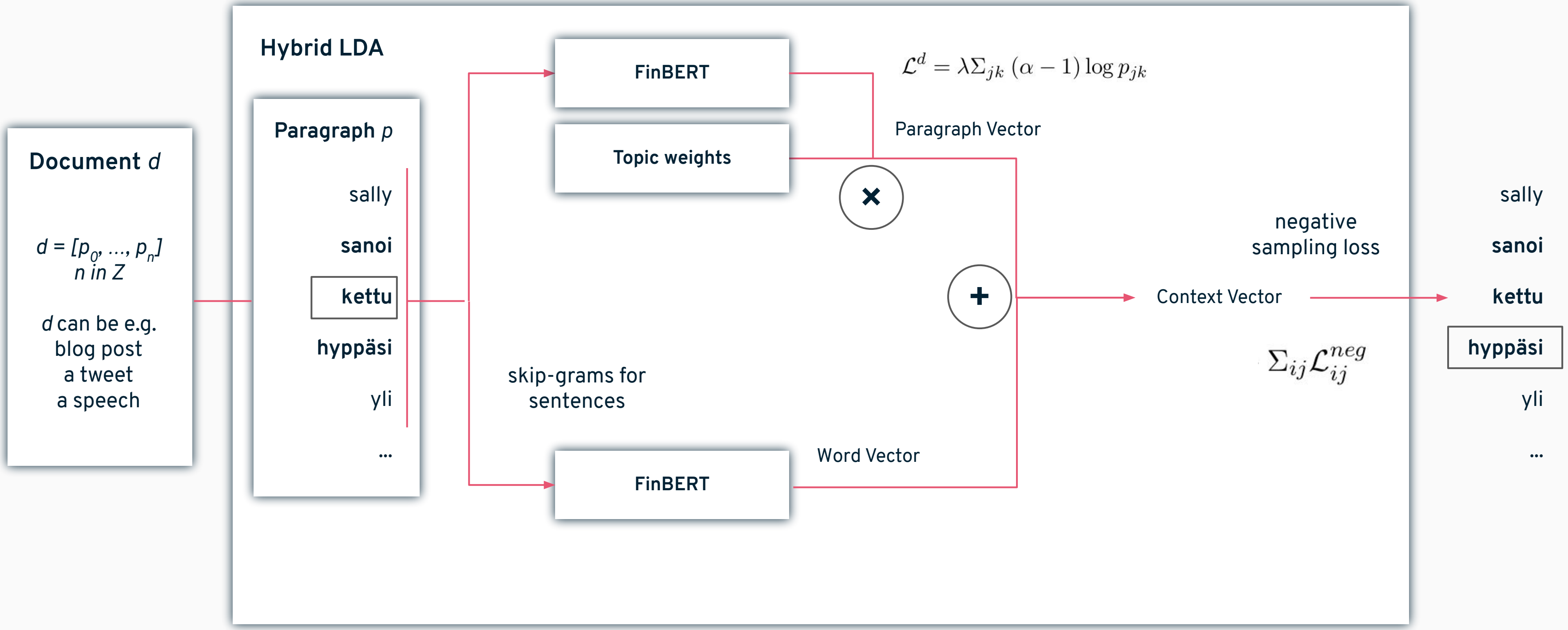
- **Online inference** to documents that have not been available during training

- Results should be **easy to interpret**

Multi Source Hybrid LDA

- Topic model that
  - leverages the generalisation and context awareness of transformers
  - provides easily interpretable sparse topic mixtures of LDA

Multi-Source Hybrid LDA Preliminary Architecture



Hybrid LDA w.r.t. *lda2vec*

- **similar** to *lda2vec* [1]:
  - loss function considers documents (paragraphs) and sampled words [3]
  - word, paragraph, context, and topic vectors are in the same space $\mathbf{R}^n$

$$\mathcal{L} = \mathcal{L}^d + \Sigma_{ij}\mathcal{L}_{ij}^{neg}$$

- **different** to *lda2vec:*
  - dense word embedding are not learned but instead FinBERT [2] is utilized to create the word and document vectors
  - possible to do (online) inference to documents outside the training set

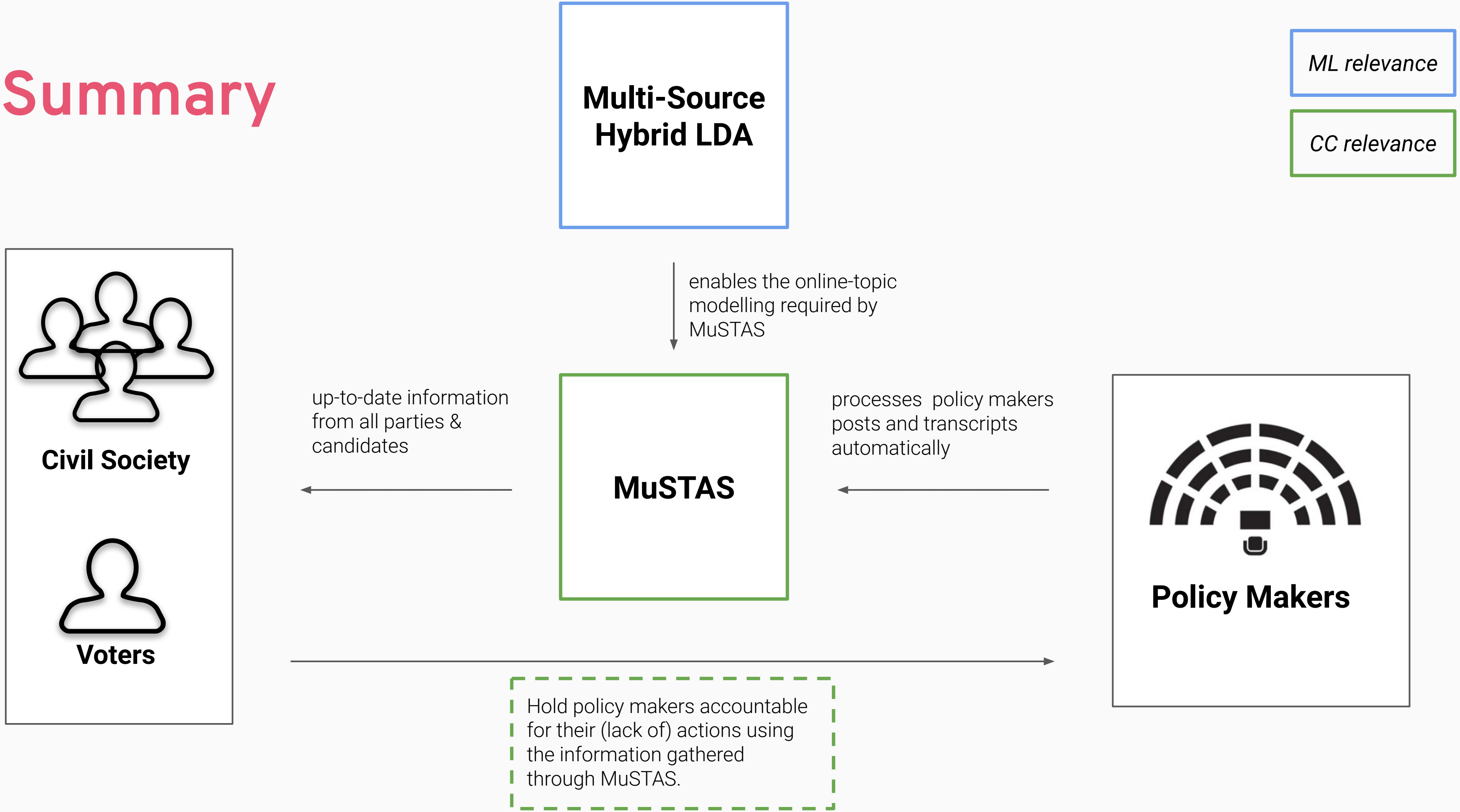
[1] Moody, C. E. Mixing Dirichlet Topic Models and Word Embeddings to Make lda2vec. May 2016

[2] Virtanen, A., Kanerva, J., Ilo, R., Luoma, J., Luotolahti, J., Salakoski, T., Ginter, F., and Pyysalo, S. Multilingual is not enough: BERT for Finnish.

[3] Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. NIPS, pages 3111–3119.

3. Summary & Next Steps

Summary



Multi-Source Hybrid LDA
ML relevance
CC relevance
enables the online-topic modelling required by MuSTAS
Civil Society
Voters
up-to-date information from all parties & candidates
MuSTAS
processes policy makers posts and transcripts automatically
Policy Makers
Hold policy makers accountable for their (lack of) actions using the information gathered through MuSTAS.

Next Steps to Implementation

If you....

- have ideas/criticism
- want to get involved
- could utilise MuSTAS
- know of grants/sponsors
- have resources or relevant literature

Help us in holding politicians accountable!

Vili Hätönen

vili@emblica.com

Fiona Melzer

F.melzer@sms.ed.ac.uk

Please get in touch!