

Representing Political Topics with Sentence Transformers

- Transfer Learning with Topic Centroids

Moritz Laurer
Centre for European Policy Studies

The ambition – A Political Discourse Classifier

(1) New machine learning and NLP techniques promise high accuracy for classifying texts in pre-defined categories. (2) Large annotated corpora of political texts are available.

➔ *Is it possible to build an automatic political discourse classifier, which identifies political ideas in diverse types of texts?*

Dataset: Manifesto Corpus

- 130.000+ English quasi-sentences from party manifestos (countries: NZ, GB, IE, AU, CA, US, ZA)
- Annotated in 7 overarching domains (e.g. economy, external relations) and 56 sub-domains (e.g. free market, anti-imperialism)

Issue: Noisy data

- Labeling scheme too complex and overlapping.
 - Each party manifesto only annotated by one annotator
- => Misclassifications & low reliability (Mikhaylov et al. 2012)

Existing Research: Mostly Softmax Classifiers

- **Best models:** Softmax-based text classifiers like CNNs or BERT models (Terechshenko et al. 2020, Chatsiou 2020, Abercrombie et al. 2019).
- **Metrics on 56 classes:** Abercrombie et al.: 0.42 F1 macro; Bilbao-Jayo and Almeida (2018): 0.42 F1;
- **Metrics on 44 classes:** Osnabrügge et al. (forthcoming) 0.42 F1 macro and 0.39 balanced accuracy (using logistic regression).

Centroid Classification with Sentence Transformers

- Embed each sentence to a dense vector with a sentence transformer
 - Calculate the centroid of all embeddings of the same class
 - Non-training sentences which are close to a centroid probably belong to the centroid's class
- ➔ Assumption: These 'topic centroids' represent the 'average meaning' of the class

Table 1. In-Domain Performance

N_classes	F1_macro	F1_micro	Accuracy_balanced
56-single	0.35	0.45	0.40
56-multi-3-thresh	0.56	0.66	0.61

Advantages for Transfer Learning

Scenario: Applying classifier to a large news article corpus

Issue: different label distribution between source and target

- Most news sentences don't belong to any category (out-of-distribution, OOD)
- News sentences can belong to more than one class, e.g. peace & internationalism (multi-label)

Topic centroids

- Sentences below threshold can be discarded as OOD
 - Sentences can be close to multiple topic centroids
- ➔ Multi-label classification with threshold increases accuracy

Work in Progress:

- Manually annotate out-of-domain data to validate assumptions for transfer learning

Caveats and Open Questions:

- In-domain metrics strongly depend on pre-processing decisions (N classes, text processing, sample balance)
- Best threshold for multi-label is unclear. Easy to overestimate performance
- Noisy data problem not solved

Table 2. Positive and Negative Examples

Text	Label_centroid	Similarity	Label_gold
a carbon tax levied on generators of mains-supplied electricity or gas;	Environmental Protection	0,72	Environmental Protection
a carbon tax levied on generators of mains-supplied electricity or gas;	Anti-Growth Economy and Sustainability	0,71	Environmental Protection
a carbon tax levied on generators of mains-supplied electricity or gas;	Market Regulation	0,70	Environmental Protection
Negotiated agreements to reduce trade barriers and expand access to foreign markets.	Protectionism: Negative	0,80	Free Market Economy
Negotiated agreements to reduce trade barriers and expand access to foreign markets.	Free Market Economy	0,80	Free Market Economy
Negotiated agreements to reduce trade barriers and expand access to foreign markets.	Protectionism: Positive	0,78	Free Market Economy