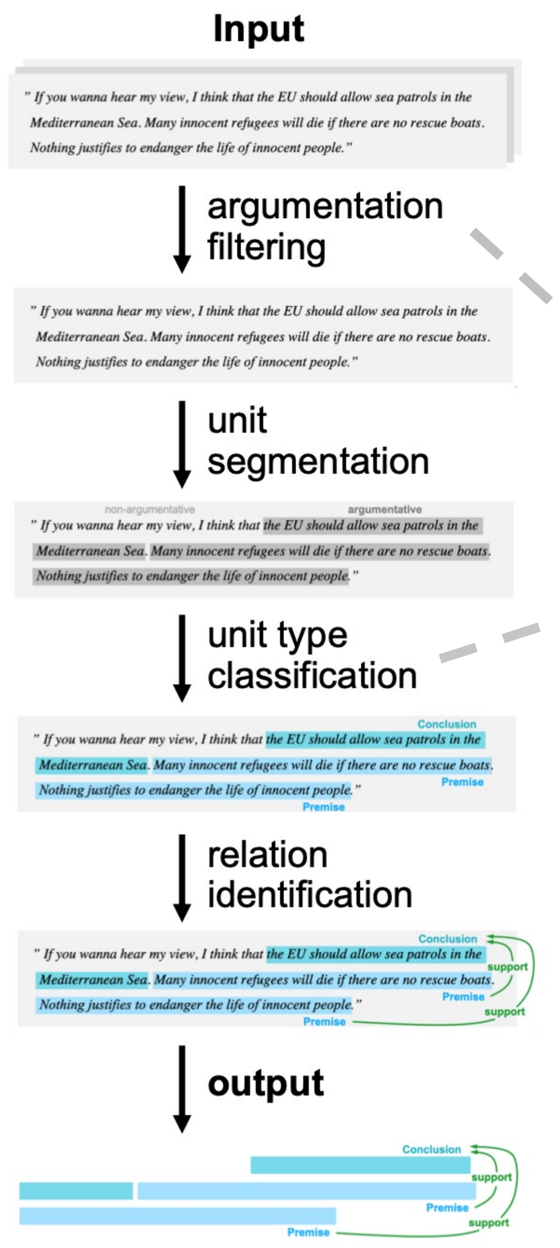
Mining Arguments in Political Domain with Transformer Language Models

Mariia Poiaganova poiaganova@uni-potsdam.de
Supervisor Prof. Dr. Manfred Stede

1 Introduction



Political discourse offers an opportunity to compare speakers' **positions on controversial topics**

Within the typical **argument mining process**, we focus on the two tasks:

Task 1: determine whether given text contains an argument or not

Task 2: among argumentative units, classify whether a unit is a claim or a premise

Research Directions

D1: Compare performance of transfer learning models to other previously reported models

D2: Study models' performance on task 2 given sentence level vs argumentative discourse unit (ADU) level sequences as input

D3: Investigate models' prediction potential on a different dataset within political domain

2.1 Data: Presidential Debates

Dataset USElecDeb60To16 v.01
(https://github.com/ElecDeb60To16/Dataset)

- Debates held among the leading candidates for the presidential and vice-presidential nominations in the US
- 39 speeches, from Kennedy and Nixon in 1960 until those between Clinton and Trump in 2016

Total	Non-Arguments	Arguments	Premises	Claims
29.621	7.252	22.280	10.316	11.964

Table 1. USElecDeb corpus statistics (per number of sentences)

Example of premises and claims

R. Nixon: "Senator Kennedy's position and mine completely different on this. I favor the present depletion allowance. I favor it not because I want to make a lot of oil men rich, but because I want to make America rich".

2.2 Data: UC-UNSC corpus creation

- We create argument structure annotations on The United Nations Security Council speeches, covering conflict in Ukraine from 2014 to 2018
- *Ukraine Conflict – UNSC (UC-UNSC)*
- 144 speeches in total from representatives of 24 different countries

Total	Non-Arguments	Claims	Premises
4.751	937	2.077	1.737

Table 2. UC-UNSC corpus statistics (per number of sentences)

Example of premises and claims

Representative of Australia, 2014: "Ukraine is living up to its Geneva commitments. It has submitted to Parliament a draft law on amnesty for protesters who surrender their weapons. It has initiated a process of constitutional reform aimed at decentralizing power".

3 Experitental Setting

We solve both Task 1 and Task 2 as **binary classification problems** and do so

- On a sentence level (if a sentence contains both a claim and a premise, it is labelled according to the longer component)
- On ADU level for the task 2 (original labelled components in both corpora)

We make the use of the following **transfer learning models**:

- BERT, RoBERTa and RoBERTa Argument (trained on ~25k manually annotated sentences of controversial topics) (Stab et al., 2018)

4 Results

- Improved performance on both tasks
- More accurate when predicting non-argumentative class on task 1
- Overall increase of 4% on RoBERTa, task 2, claim vs premise detection
- We observe that RoBERTa is slightly more robust compared to BERT

Task	Majority Baseline	LSTM (former implementation)	BERT	RoBERTa
Task 1	Arg 0.810 None 0.00 Avg 0.551	0.902 0.524 0.818	0.907 0.617 0.842	0.912 0.613 **0.846**
Task 2	Claim 0.680 Premise 0.00 Avg 0.350	0.685 0.654 0.656	0.723 0.625 0.675	0.711 0.675 **0.693**

Table 3. F1-scores for sentence-level classification tasks on the USElecDeb test set

Task	Majority Baseline	BERT	RoBERTa
Task 2	Claim 0.715 Premise 0.00 Avg 0.398	0.761 0.682 **0.726**	0.758 0.690 **0.728**

Table 4. F1-scores for ADU-level classification tasks on the USElecDeb test set

- Both models show better distinguishing between premises and claims on the ADU level compared to sentence level
- A single sentence can contain both a premise and a claim; segmenting such input sequences is important

- Comparable performance on the original test and new UC-UNSC corpus on task 2; but worse performance on task 1 ->
- Conceptualization of arguments might be different in the two corpora

Task	Majority Baseline	BERT
Task 1	Arg 0.891 None 0.00 Avg 0.715	0.906 0.484 **0.823**
Task 2	Claim 0.705 Premise 0.00 Avg 0.390	0.723 0.707 **0.716**

Table 5. F1-scores for sentence-level classification tasks on test sets, trained on full USElecDeb, tested on UC-UNSC

5 Future Directions

- More annotators to provide more reliable annotations, IAA
- Annotating and predicting claim- premise relations (support/attack)

References

Project's repository:
https://github.com/a-moi/political-argument-mining

Haddadan, S., Cabrio, E., & Villata, S. (2019). Yes, we can! Mining Arguments in 50 Years of US Presidential Campaign Debates. *ACL*.
Christian Stab, Tristan Miller, Benjamin Schiller, Pranav Rai, and Iryna Gurevych. 2018. Cross-topic Argument Mining from Heterogeneous Sources. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3664–3674, Brussels, Belgium. Association for Computational Linguistics
Argument Mining Process graph, author: Henning Wachsmuth (https://en.cs.uni-paderborn.de/fileadmin/informatik/fg/css/computational-argumentation-s20/ca20-part05-mining.pdf)