Experiments on a Novel Approach for the Detection of Propaganda Techniques in News Articles

SemEval-2020 Task 11 as part of the NLP course at ETH

18/01/22

Agenda

I. The task at hand
II. Our approach
III. Its shortcomings, and how we overcame them
IV. Our results
V. An error analysis
VI. Outro

The task

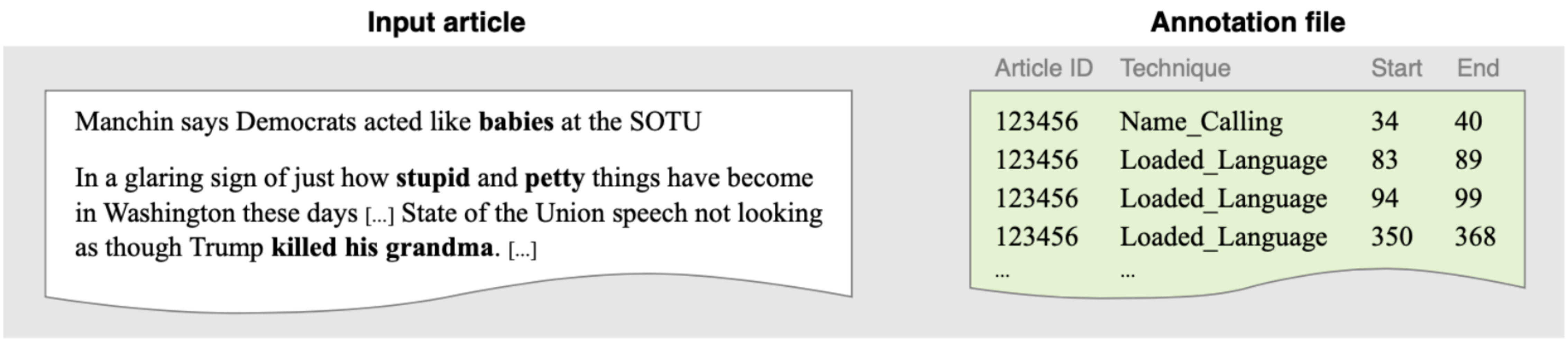
Detecting propaganda spans

- Given input articles, the task is divided in two:

  1. Span Identification (SI)

  2. Technique Classification (TC)

- There are 14 classes of propaganda



Input data and annotation, visualised[1].

(1) Da San Martino, Barrón-Cedeño, Wachsmuth, Petrov, and Nakov. 2020a. Semeval-2020 task 11: Detection of propaganda techniques in news articles

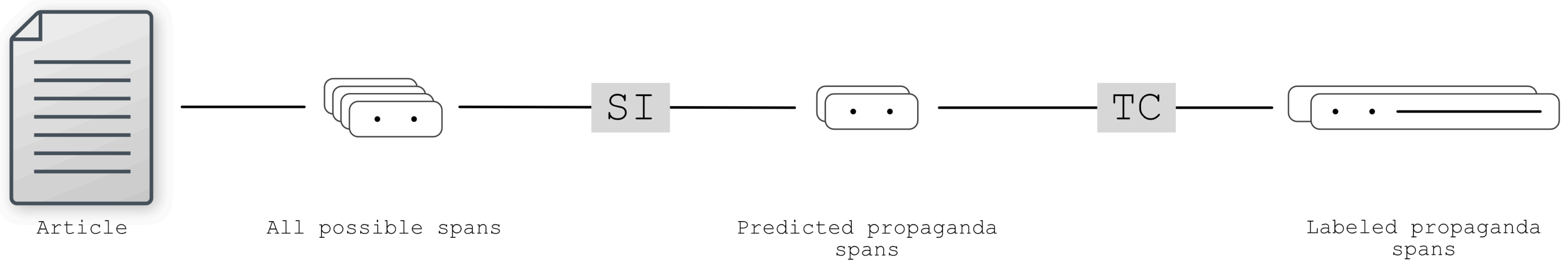
Our approach

- Not wanting TC to have to rely on a perfect set of spans, we introduced two major changes:

    1. TC will train on the set of spans predicted by SI (we enrich our gold spans)

    2. A 15th class, “*Not Propaganda*” is introduced for TC

- Relies on new assumption

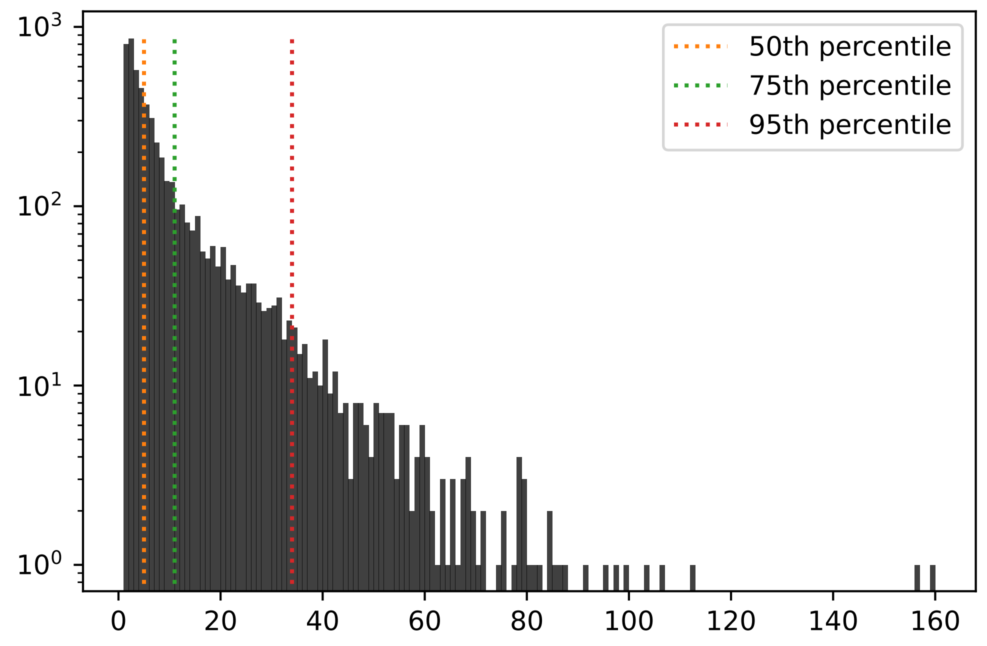


General proposed system architecture.

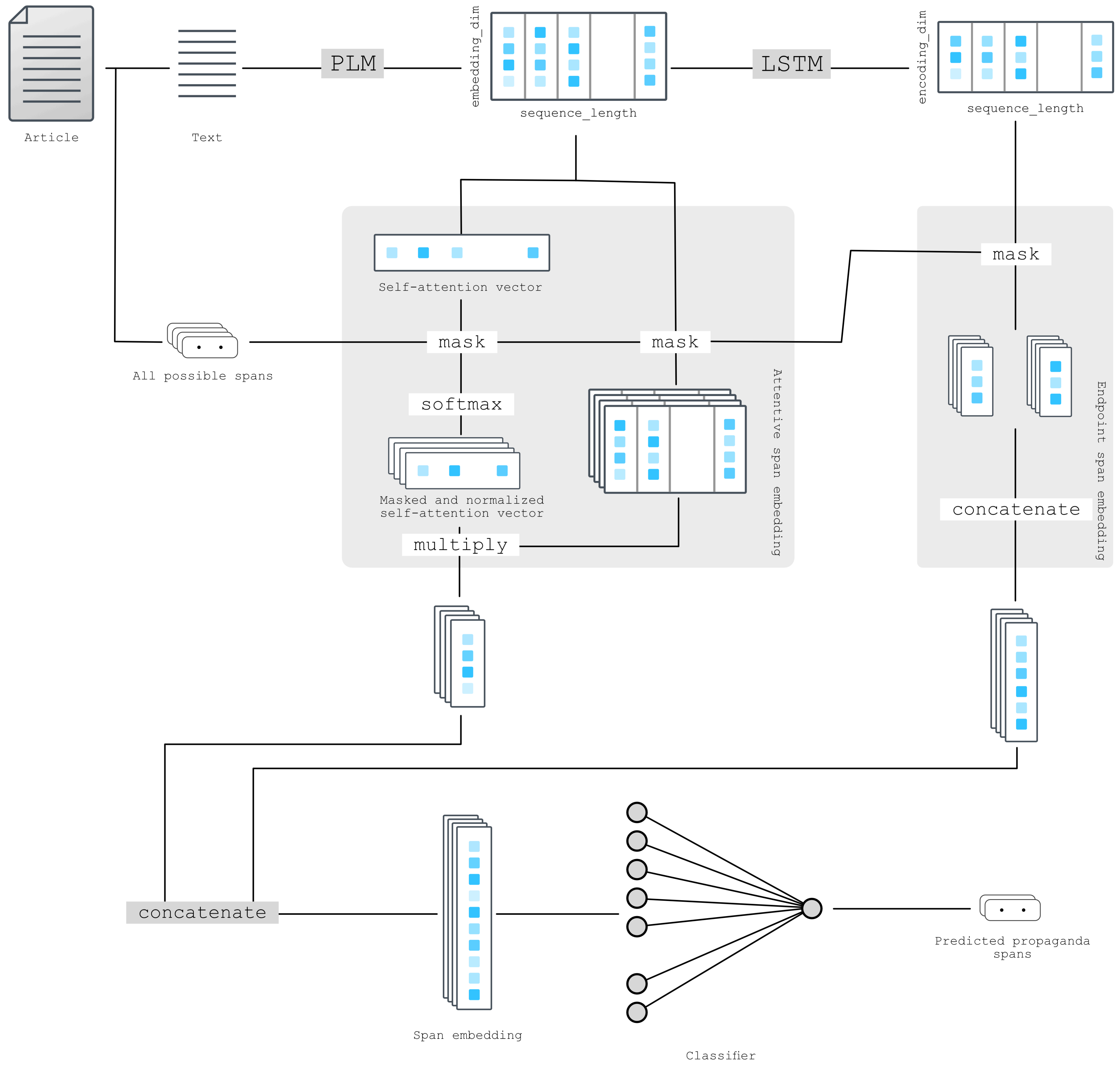
Our approach

The SI model

- Span classification instead of sequence labelling
- Some preprocessing on the text



- Embedding and classifying, with weights

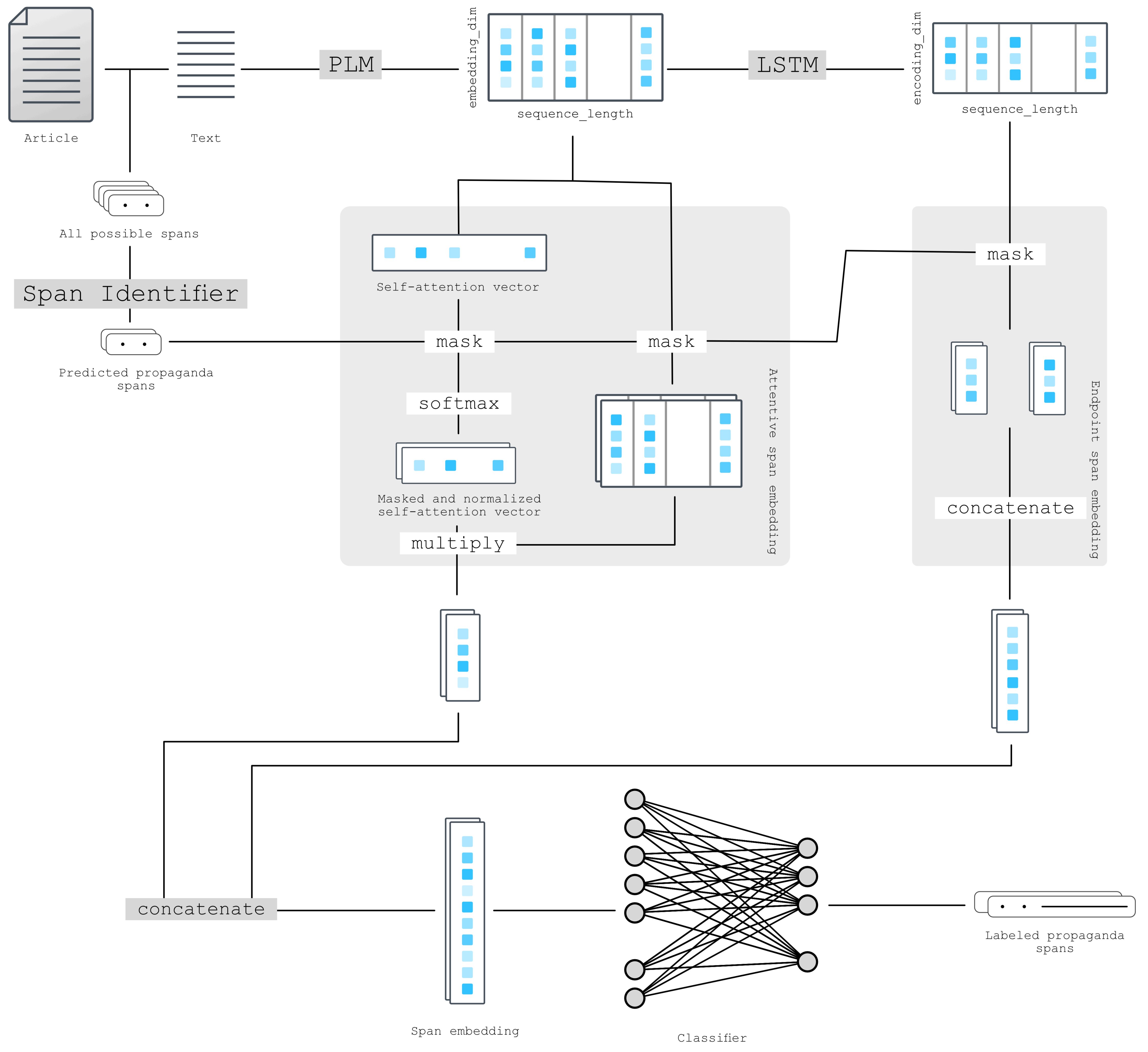


SI architecture.

Our approach

The TC model

- SI as additional pruning
- Adapting the data
- Embedding and classifying, again, with weights



TC architecture.

Shortcomings

The SI model

- Exponential spans considered
- Computationally very expensive
- Unable to train on cluster's GPUs
- Unable to perform fine-tuning

$$\sum_{x=1}^{20} \binom{100}{x} < \binom{100}{20} \approx \mathbf{5.35e{+}20}$$

$$\sum_{x=1}^{10} \binom{100}{x} < \binom{100}{10} \approx \mathbf{1.73e{+}13}$$

Shortcomings

The TC model

- SI spans were not perfect matches
- Even weights couldn’t resolve that
- Explored partially overlapping spans
- Our approach did not work in practice

Threshold	1	≥ 0.5	≥ 0.25	> 0
Percentage	0.041	0.205	0.301	0.397

Percentages of predicted spans which match different values of IoU score.

New solution

Alternative TC model

- We looked at the solution proposed by SemEval
- Gold spans from perfect dataset
- Removal of our 15th class

Results

- Ranked 8/45 teams for SI
- Achieved F1 0.57572 for TC

- Only 10 and 1 epochs resp. !
- No hyperparameter fine-tuning!

Model	Custom F_1	Precision	Recall
BERT	0.40008	0.29371	0.62722
RoBERTa	0.42649	0.32754	0.61107
XLNet	0.37930	0.26213	0.68590

Model results on SI task with validation data.

Model	Custom F_1	Precision	Recall
BERT	0.29651	0.17528	0.96147
RoBERTa	0.46072	0.40635	0.53189
XLNet	0.43133	0.50394	0.37701

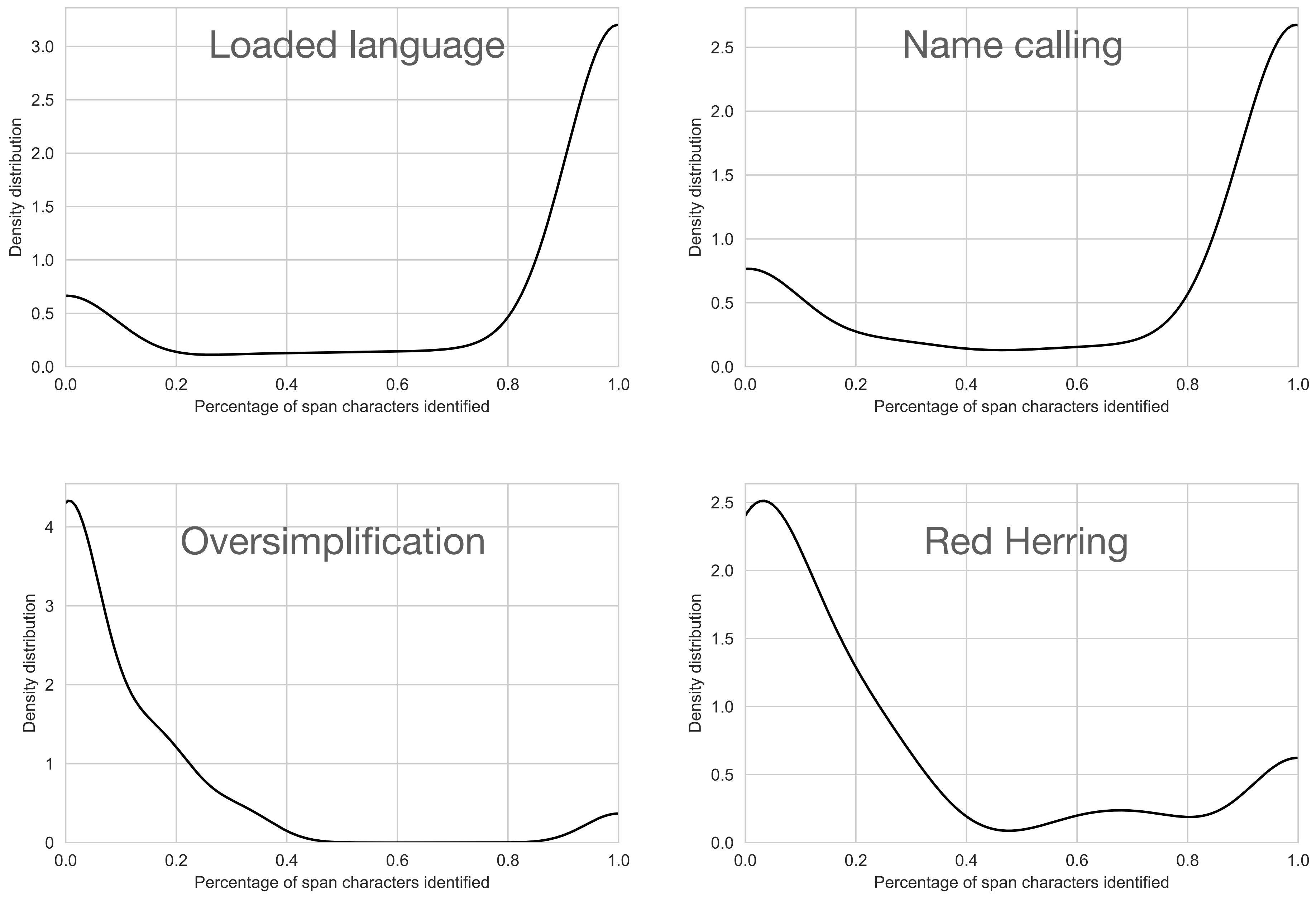
Model results on SI task with test data.

Error analysis

The SI model

	Loaded Language	Name Calling	Repetition	Flag Waving	Exaggeration	Doubt	Prejudice	Slogans	Red Herring	Appeal to Authority	Reductio ad hitlerum	Oversimplification	Cliches	Authority	Total
Not identified	51	35	56	18	28	42	9	10	17	7	4	13	9	7	306
Partially identified	23	18	3	19	12	18	17	4	8	1	1	4	3	1	132
Totally identified	251	130	86	50	28	6	18	26	4	6	0	1	5	6	617
Total	325	183	145	87	68	66	44	40	29	14	5	18	17	14	1055

SI results broken down by propaganda technique. In this setting, a gold span was considered *totally identified* if at least 75% of its characters were labeled as propaganda, *partially identified* if a percentage between 15% and 75% of its characters were labeled as propaganda, *not identified* if less than 15% of its characters were labeled as propaganda.



Distribution of identification percentage of gold spans which belong to four different propaganda technique. It can be observed how less frequent techniques in the training set (Figures 5a and 5b) are much harder to label compared to more frequent techniques (Figures 5c and 5d).

Error analysis

The TC model

- Error analysis for original TC model not valuable
- Instead, classification results from alternative TC model were investigated

	Loaded	Labeling	Repetition	Exaggeration	Doubt	Prejudice	Flag	Oversimplification	Slogans	Authority	Fallacy	Cliches	Whataboutism
Loaded	0.715	0.059	0.105	0.063	0.008	0.025	0.000	0.000	0.004	0.013	0.004	0.004	0.000
Labeling	0.098	0.591	0.197	0.015	0.015	0.030	0.015	0.015	0.000	0.008	0.000	0.008	0.008
Repetition	0.182	0.136	0.636	0.000	0.045	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Exaggeration	0.333	0.000	0.200	0.467	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Doubt	0.074	0.000	0.296	0.000	0.444	0.000	0.111	0.000	0.000	0.000	0.000	0.074	0.000
Prejudice	0.214	0.000	0.000	0.000	0.071	0.357	0.000	0.071	0.000	0.214	0.000	0.071	0.000
Flag	0.000	0.000	0.040	0.000	0.000	0.040	0.840	0.040	0.000	0.040	0.000	0.000	0.000
Oversimplification	0.000	0.077	0.154	0.000	0.000	0.000	0.077	0.692	0.000	0.000	0.000	0.000	0.000
Slogans	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000
Authority	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Fallacy	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000
Cliches	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Whataboutism	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Normalised confusion matrix obtained from results of alternative TC. Rows represent the correct labels and columns the predicted ones.

Future work

- Fine-tuning hyperparameters
- Exploration of add-on features
- Using data augmentation techniques
- Improving top-layer classification
- Exploring more PLMs…

Outro

- We checked the ETH Zürich NLP lecture notes with our system
- But also our paper itself..!

- highly inefficient → loaded language
- the cheap gradient → name calling
- how does papa eat caviar? → doubt
- everybody loves someone else → slogan
- state-of-the-art results → exaggeration
- “not propaganda” → slogan

Thank you for your attention!