

Experiments in Detecting Persuasion Techniques in the News

Seunghak Yu¹ Giovanni Da San Martino² Preslav Nakov²

¹MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, USA

²Qatar Computing Research Institute, HBKU, Qatar

seunghak@csail.mit.edu {gmartino, pnakov}@hbku.edu.qa

Computational Propaganda

- Propaganda has been widely used since the advent of mass media



- However, Internet and social media “have allowed cross-border computational propaganda by foreign states or even private organizations” (Bolsover and Howard, Big Data 2017)
- Can we automatically detect the use of propaganda?
Can we make America (and the world) aware again?
- Current approaches provide document-level predictions
 - rely on gold labels based on distant supervision → noisy
 - lack model explainability

Propaganda Techniques

- Propaganda is conveyed through a series of rhetorical and psychological techniques



Figure 1: Bandwagon: join in because everyone else is taking the same action.



Figure 2: Name calling.

Greta Thunberg: “We are in the middle of the sixth mass extinction, with more than 200 species getting extinct every day.”

Figure 3: Appeal to fear.

Propaganda Techniques Corpus

450 news articles from 48 sources (21,230 sentences, 350K tokens) annotated at the fragment level with 18 propaganda techniques.

Stereotyping name_calling or labeling			
1	Manchin says Democrats acted like babies	at the SOTU	
2	Democrat West Virginia Sen. Joe Manchin says his colleagues’ refusal to stand or applaud during President Donald Trump’s State of the Union speech was disrespectful and a signal that		
<i>(Black-and-white Fallacy)</i>			
	the party is more concerned with obstruction than it is with progress.		
<i>(Loaded language)</i>			
4	In a glaring sign of just how stupid and petty things have become in Washington these days, Manchin was invited on Fox News Tuesday morning to discuss how he was one of the only Democrats in the chamber for the State of the Union speech		
	<i>(Exaggeration)</i>	<i>(Loaded language)</i>	
	not looking as though Trump killed his grandma.		
6	As Manchin noted, many Democrats bolted as soon as Trump’s speech ended in an apparent effort to signal		
	<i>(Exaggeration)</i>		
	they can’t even stomach being in the same room as the president		

Annotation Process

- Phase 1: two annotators, a_i and a_j , independently annotated the same article
- Phase 2: a_i and a_j discussed with a consolidator c_1 all instances to come up with a final annotation.

The table shows γ inter-annotator agreement for spans only and spans + labels between two annotators and one annotator and one consolidator.

Annotations		spans (γ_s) +labels (γ_{sl})	
a_1	a_2	0.30	0.24
a_3	a_4	0.34	0.28
a_1	c_1	0.58	0.54
a_2	c_1	0.74	0.72
a_3	c_2	0.76	0.74
a_4	c_2	0.42	0.39

Tasks and Evaluation Measure

- FLC:** detect the text fragments in which a propaganda technique is used and identify the technique.
Spans is a lighter version of the task in which only the span has to be identified.
- SLC** detect the sentences that contain one or more propaganda techniques (binary task).

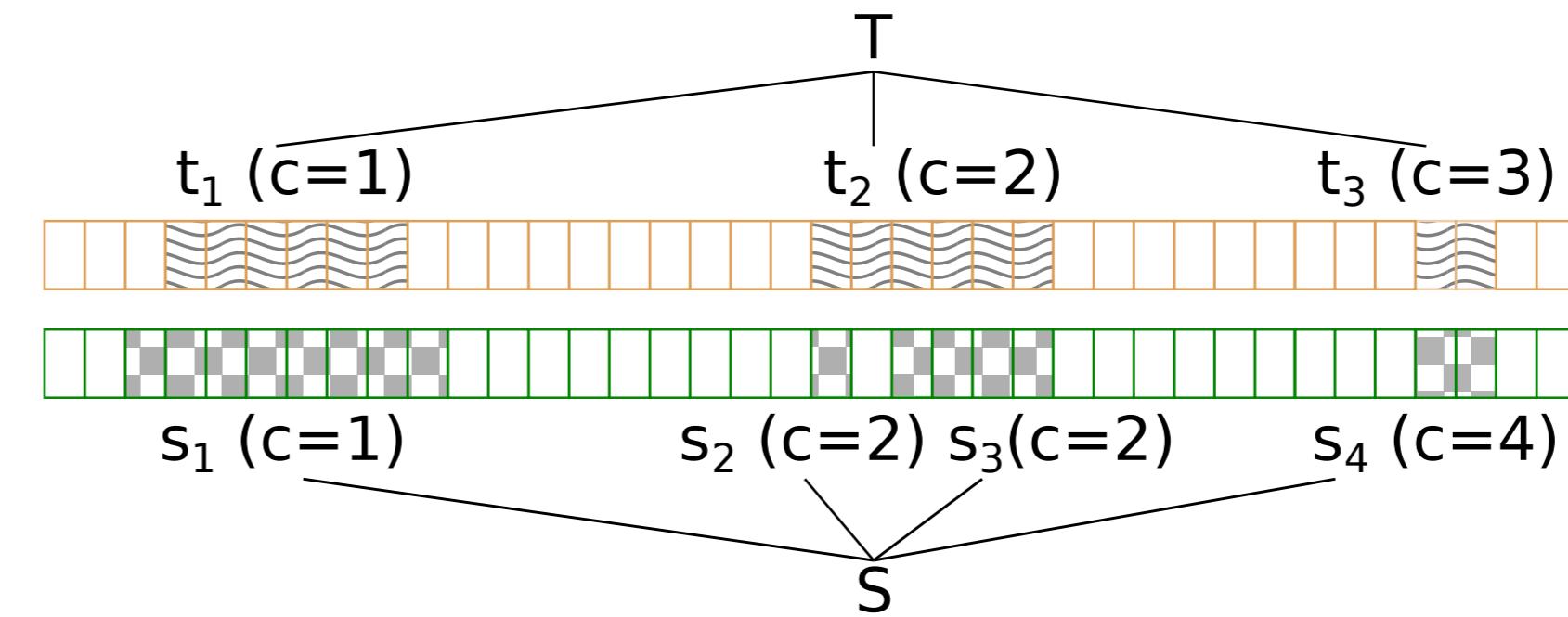
An evaluation measure for Task FLC needs to be defined. We use a variant of the standard F_1 (and Precision, P, and Recall, R) taking into account partially overlapping spans:

$$P(S, T) = \frac{1}{|S|} \sum_{s \in S, t \in T} C(s, t, |s|), \quad R(S, T) = \frac{1}{|T|} \sum_{s \in S, t \in T} C(s, t, |t|),$$

where

$$C(s, t, h) = \frac{|(s \cap t)|}{h} \delta(l(s), l(t))$$

here $\delta(a, b) = 1$ if $a = b$, and 0 otherwise.



Example of gold annotation (top) and the predictions of a supervised model (bottom) in a document represented as a sequence of characters. The class of each fragment is shown in parentheses.

Models

- Multi-Granularity Network:** It drives the higher-granularity task (FLC, g_2) on the basis of the lower-granularity information (SLC, g_1) through a trainable gate f :

$$o_{g_2} = f(o_{g_1}) * o_{g_2}$$

and we used a weighted sum of losses with a hyper-parameter α

$$L_J = L_{g_1} * \alpha + L_{g_2} * (1 - \alpha)$$

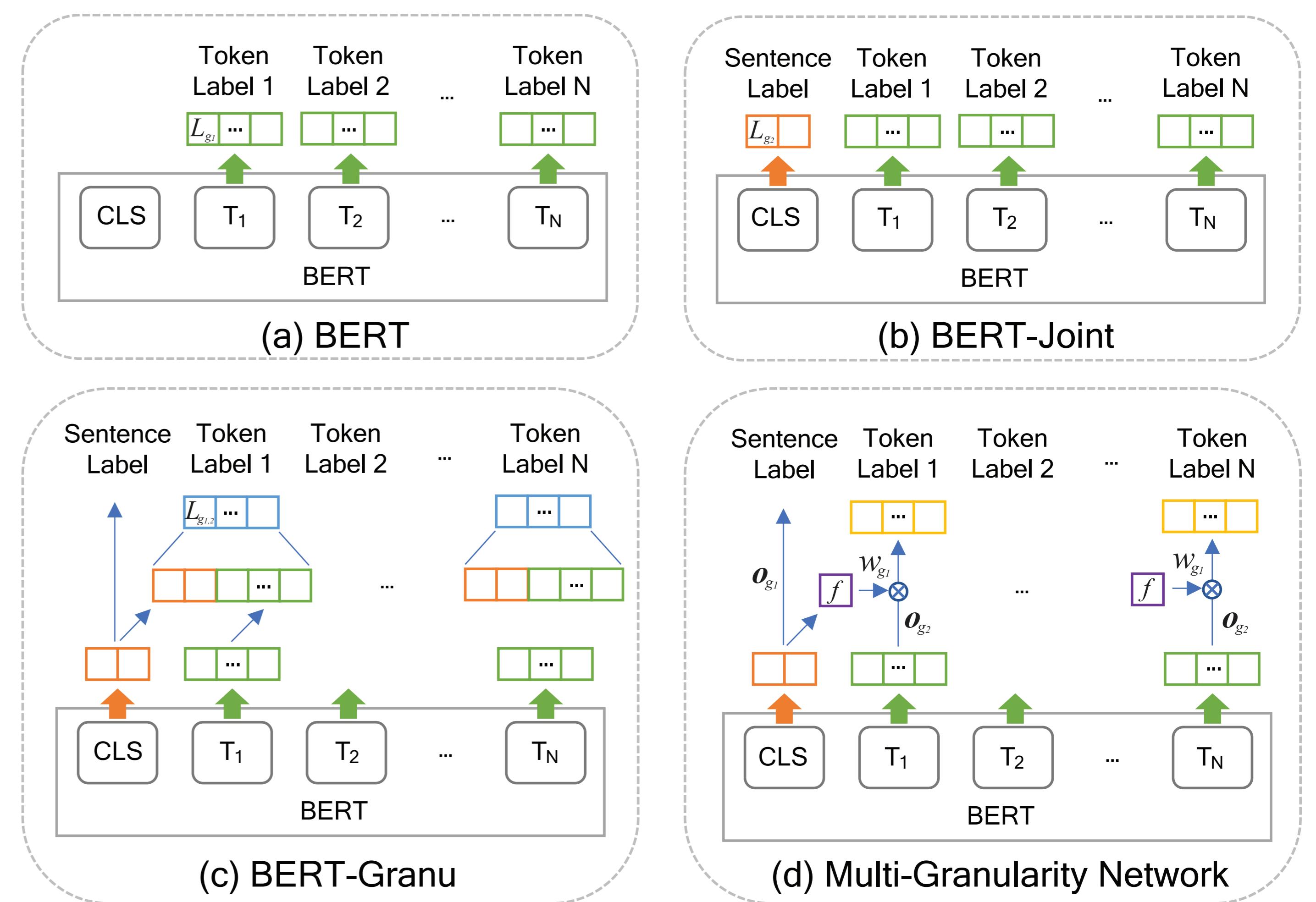


Figure 4: The architecture of the baseline models (a-c), and of our proposed multi-granularity network (d).

Experiments

Model	Spans			FLC Task			SLC Task		
	P	R	F_1	P	R	F_1	P	R	F_1
BERT	39.57	36.42	37.90	21.48	21.39	21.39	63.20	53.16	57.74
Joint	39.26	35.48	37.25	20.11	19.74	19.92	62.84	55.46	58.91
Granu	43.08	33.98	37.93	23.85	20.14	21.80	62.80	55.24	58.76
Multi-Granularity									
ReLU	43.29	34.74	38.28	23.98	20.33	21.82	60.41	61.58	60.98
Sigmoid	44.12	35.01	38.98	24.42	21.05	38.98	62.27	59.56	60.71

Table 1: Evaluation of the models for Spans, FLC and SLC tasks. The proposed models improve over the baselines.

Conclusion and Future Work

- Our fine-grained task can complement document-level judgments, both to come out with an aggregated decision and to explain why it has been flagged as potentially propagandistic.
- We plan to build an online platform to annotate propaganda techniques and expand the corpus.

What We Are Up To

- SemEval 2020 Task 11 on Fine Grained Propaganda Detection: <https://propaganda.qcri.org/semeval2020-task11>
- Our Propaganda Analysis Project (where you can find this paper): <https://propaganda.qcri.org>
- The Tanbih Project, which aims to limit the effect of “fake news”, propaganda and media bias by making users aware of what they are reading: <http://tanbih.qcri.org>