

POLITICS: Pretraining with Same-story Article Comparison for Ideology Prediction and Stance Detection

Yujian Liu^{1,*}, Xinliang Frederick Zhang^{1,*}, David Wegsman¹,
Nick Beauchamp², and Lu Wang¹

¹University of Michigan, ²Northeastern University
*Equal Contributions



Background

Ideology Prediction:

Conservatives continue to support **baseless** claims of election fraud ... Trump isn't the only Republican making **unsubstantiated** claims about voter fraud, a move that may put lives at risk.

Political ideology?

Liberal

Stance Detection:

I don't **believe** life begins at conception--I **know** life begins at conception. #nrlc2015

Stance towards legal abortion?

Against

News Story: Donald Trump tests positive for COVID-19.

Daily Kos (left): It's now clear that Donald Trump **lied** to the nation about when he received a positive test for COVID-19. ... they're continuing to act as if nothing has changed—and that **disregarding science** and **lying** to the public are the only possible strategies.

The Washington Times (right): Trump says he's "doing very well" ... President Trump thanked the nation for supporting him Friday night as he left the White House to be hospitalized for COVID-19. "I want to thank everybody for the tremendous support. ..." Mr. Trump said in a video recorded at the White House.

Breitbart (right): President Donald Trump thanked Americans for their support on Friday as he traveled to Walter Reed Military Hospital for further care after he was diagnosed with coronavirus. "I think I'm doing very well..." Trump said in a video filmed at the White House and posted to social media.

Same-story Article Comparison.

Subjective phrases: **bold**
Objective events: *italics*

BigNews

	Daily Kos	HPO	CNN	WaPo	NYT	USA Today	AP	The Hill	TWT	FOX	Breitbart
Ideology	L	L	L	L	L	C	C	C	R	R	R
# articles	100,828	241,417	64,988	198,529	173,737	170,737	279,312	322,145	243,181	330,166	206,512
# words	738.7	729.9	655.7	803.2	599.4	691.7	572.3	426.3	522.7	773.5	483.5

- **BIGNEWS**: 3.7M US political news articles from 11 news outlets
- **BIGNEWSALIGN**: 1M clusters of articles reporting the same story

POLITICS

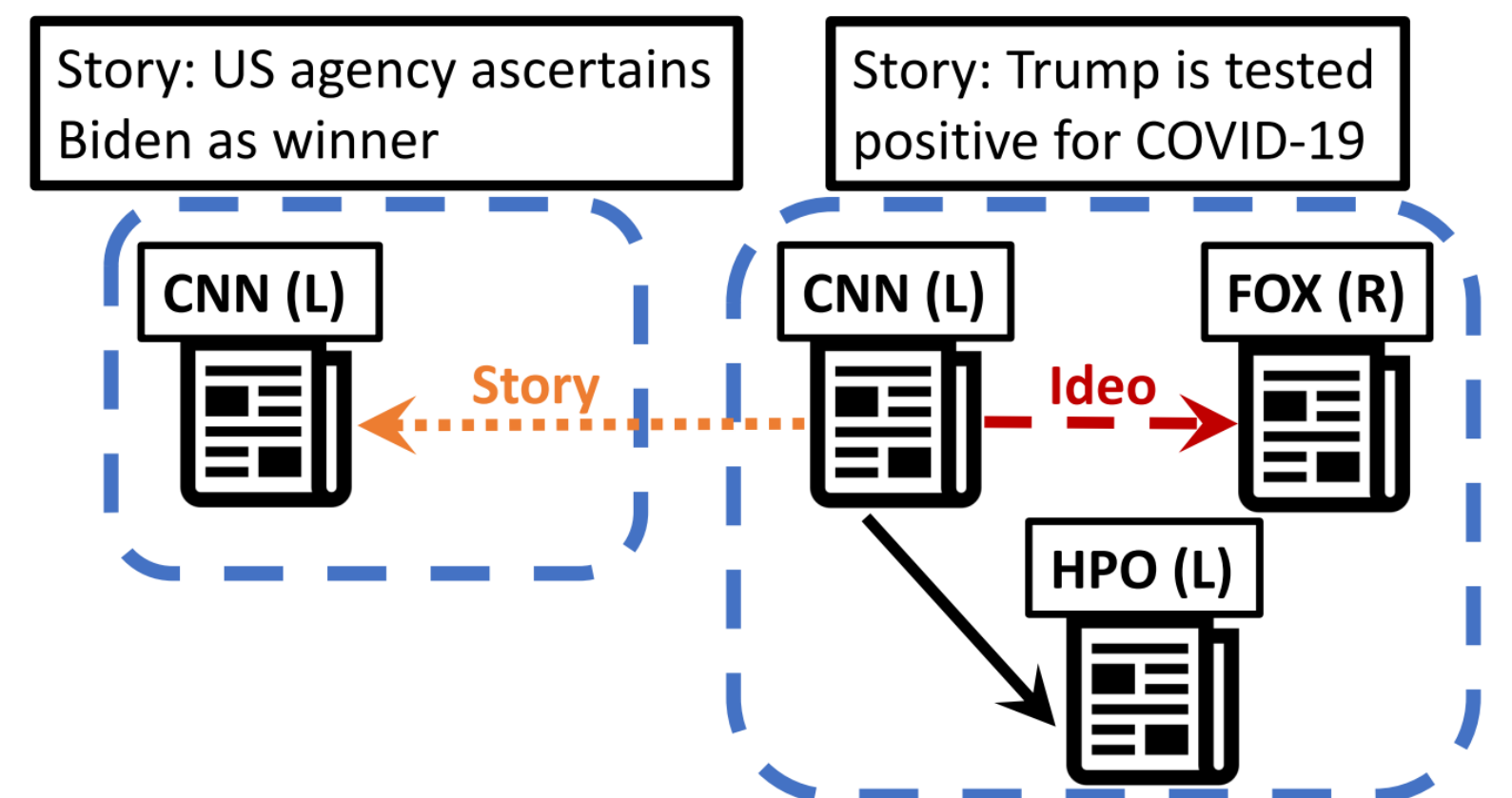
- **Ideology Objective**: Compact articles with similar ideology and contrast articles with opposite ideology

$$\mathcal{L}_{\text{ideo}} = \sum_{t \in \mathcal{T}_{\text{ideo}}} \left[\left\| \mathbf{t}^{(a)} - \mathbf{t}^{(p)} \right\|_2 - \left\| \mathbf{t}^{(a)} - \mathbf{t}^{(n)} \right\|_2 + \delta_{\text{ideo}} \right]_+$$

- **Story Objective**: Ensure the model to focus on meaningful content instead of overly relying on shortcuts

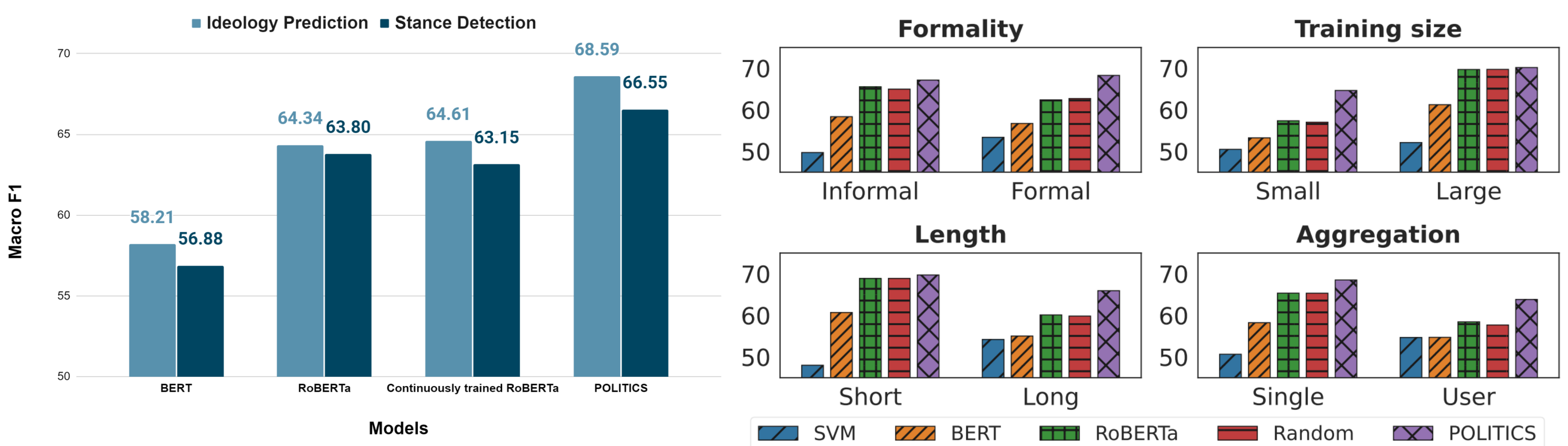
$$\mathcal{L}_{\text{story}} = \sum_{t \in \mathcal{T}_{\text{story}}} \left[\left\| \mathbf{t}^{(a)} - \mathbf{t}^{(p)} \right\|_2 - \left\| \mathbf{t}^{(a)} - \mathbf{t}^{(n)} \right\|_2 + \delta_{\text{story}} \right]_+$$

- **Entity and sentiment focused MLM objective**



Construction of the ideology and story objectives.

Results & Analysis



- **Naive MLM training does not help on evaluated tasks**
- **POLITICS performs best across the board**
- **POLITICS achieves better generalizability**

- **POLITICS is especially good at**
 - Long documents
 - Formal texts
 - Few-shot learning scenarios

Code: github.com/launchnlp/POLITICS

Contact: {yujianl, xlfzhang}@umich.edu

