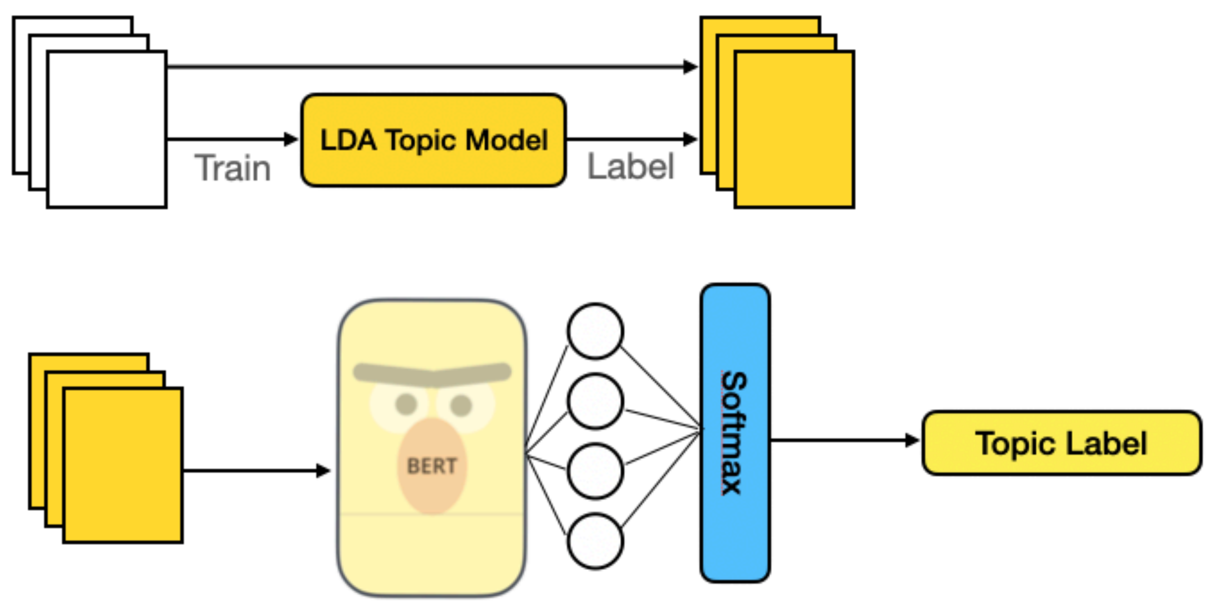
1. Motivation

- Contextualized topic models (CTMs) let us do polylingual topic modeling **without** explicit cross-lingual alignments! Model on English, then transfer.
- For other tasks, we fine-tune encoders (like BERT) on supervised data to improve performance. But **topic modeling is unsupervised**!
- **What should supervision look like for this task?**

2. Fine-tuning Tasks

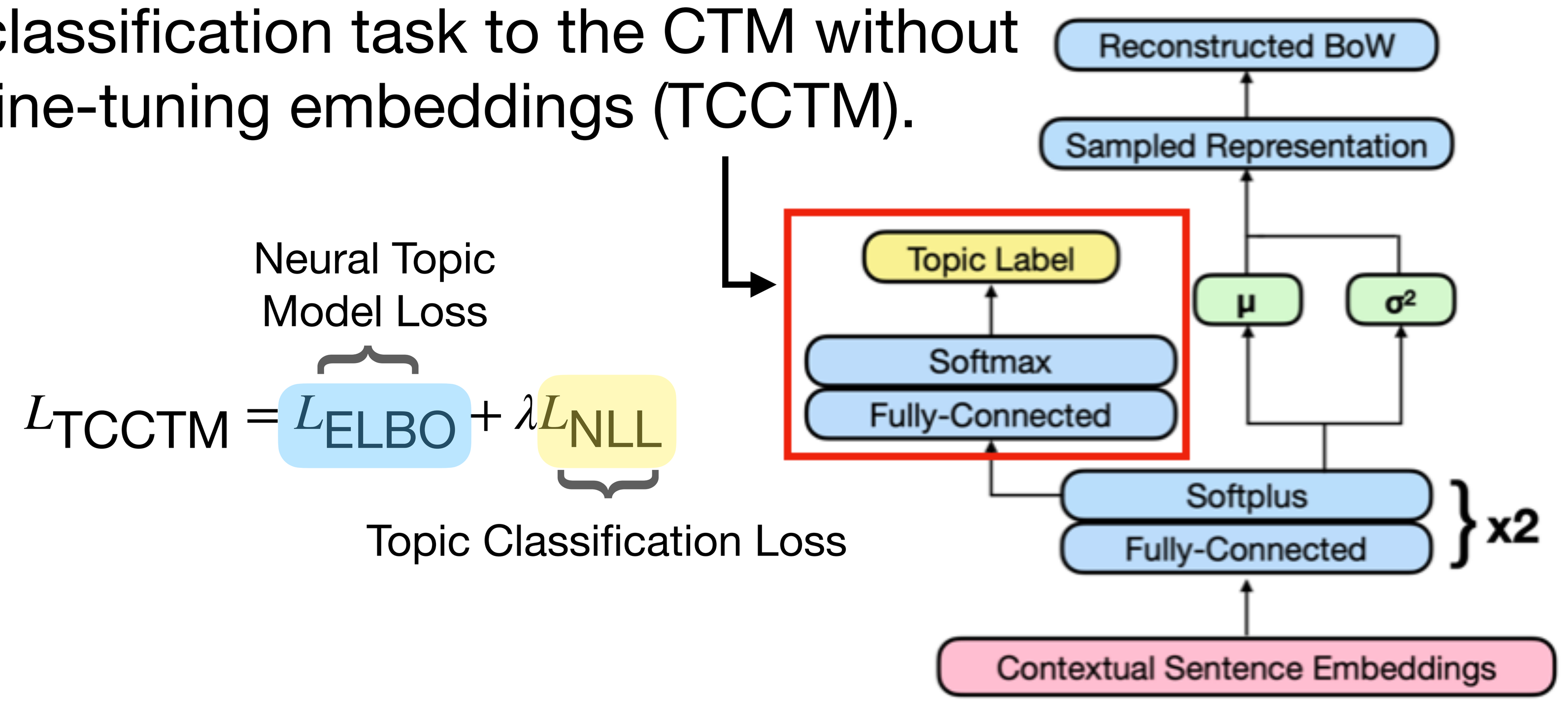
- Fine-tune mBERT and XLM-R on existing tasks: NLI (MultiNLI+SNLI), document classification (MLDoc)
- We propose **topic classification**: bootstrap supervision using only the data we topic model!



(It's just document classification with auto-generated labels.)

3. Why Does It Work?

- It works! But maybe just because the encoder sees in-domain data during fine-tuning?
- Let's do **continued pre-training** (CPT) on the topic modeling data to see if that's why.
- Let's also try adding the topic classification task to the CTM without fine-tuning embeddings (TCCTM).



Fine-tuning Encoders for Improved Monolingual and Zero-shot Polylingual Neural Topic Modeling

Aaron Mueller and **Mark Dredze**
Johns Hopkins University

4. Results

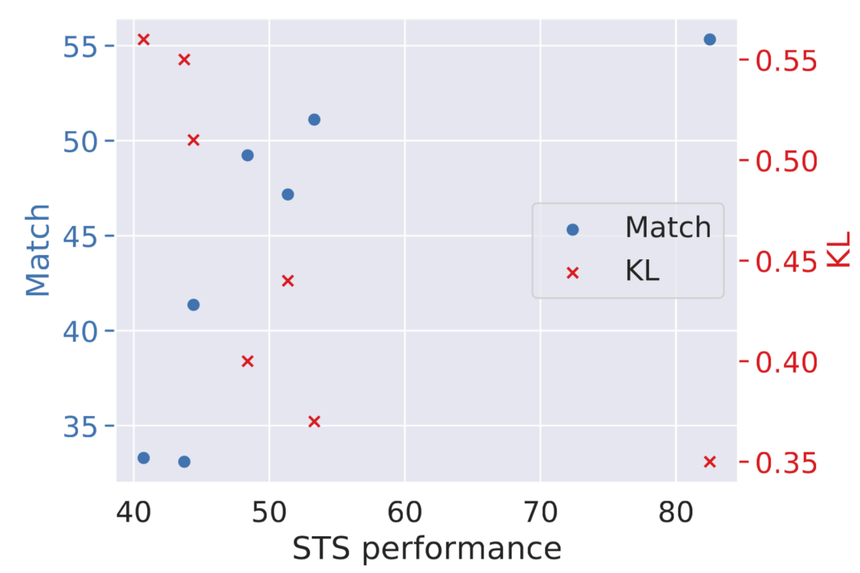
Topic classification induces most coherent topics! TCCTM is 2nd best and CPT is bad, so the topic classification objective (not in-domain data) is what helps.

Model	Fine-tuning	NPMI (XLM-R)	NPMI (mBERT)	Neural model	Fine-tuned embeddings	Topic classification	In-domain data
LDA	–		0.129				
ProdLDA	–		0.129	✓			
CTM	None	0.144	0.144	✓			
	NLI	0.153	0.152	✓	✓		
	Doc. Class.	0.156	0.153	✓	✓		
	Topic Class. (COVID)	0.156	0.153	✓	✓	✓	
	Topic Class. (Wiki)	**0.160**	**0.156**	✓	✓	✓	✓
CPT+CTM	None	0.147	0.147	✓			✓
	NLI	0.150	0.149	✓	✓		✓
	Topic Class. (COVID)	0.148	0.147	✓	✓	✓	✓
	Topic Class. (Wiki)	0.151	0.149	✓	✓	✓	✓
TCCTM	None	0.157	0.154	✓		✓	
	NLI	0.152	0.151	✓	✓	✓	
	Doc. Class.	0.153	0.152	✓	✓	✓	

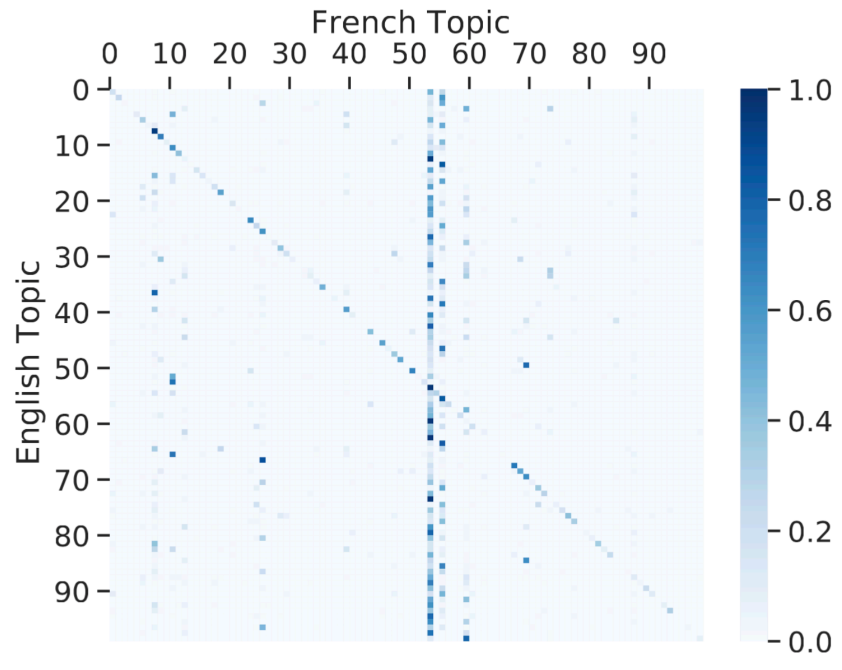
Fine-tuning on *anything* improves zero-shot topic transfer.

	French		German		Portuguese		Dutch		MEAN	
Model	Match	KL	Match	KL	Match	KL	Match	KL	Match	KL
CTM (No FT)	20.11	0.71	41.68	0.46	24.85	0.67	46.74	0.40	33.30	0.56
CTM+FT (NLI)	**53.68**	0.39	**56.29**	**0.33**	**54.38**	**0.36**	**56.98**	0.31	**55.33**	0.35
CTM+FT (DC)	35.53	0.61	42.09	0.49	38.12	0.53	49.70	0.40	41.36	0.51
CTM+FT (TC, COVID)	41.09	0.54	46.39	0.47	43.56	0.48	51.11	0.40	45.54	0.47
CTM+FT (TC, Wiki)	45.02	0.50	51.11	0.40	42.58	0.49	50.68	0.40	47.17	0.44
CPT (No FT)	23.62	0.68	40.75	0.45	22.89	0.65	45.13	0.42	33.10	0.55
CPT+FT (NLI)	43.43	0.45	48.09	0.38	43.04	0.46	49.53	0.38	46.02	0.42
CPT+FT (TC, COVID)	41.70	0.53	43.67	0.44	39.91	0.60	47.44	0.43	43.18	0.50
CPT+FT (TC, Wiki)	47.02	0.45	51.53	0.36	45.83	0.44	52.54	0.34	49.23	0.40
TCCTM (No FT)	18.81	0.71	41.18	0.46	19.21	0.72	45.49	0.42	31.17	0.58
TCCTM+FT (NLI)	53.30	**0.38**	55.52	**0.33**	53.75	0.37	56.40	**0.30**	54.74	**0.34**
TCCTM+FT (DC)	41.83	0.51	48.72	0.42	38.80	0.53	49.73	0.39	44.77	0.46
Random	0.92	1.48	1.22	1.39	1.24	1.48	1.09	1.44	1.12	1.44

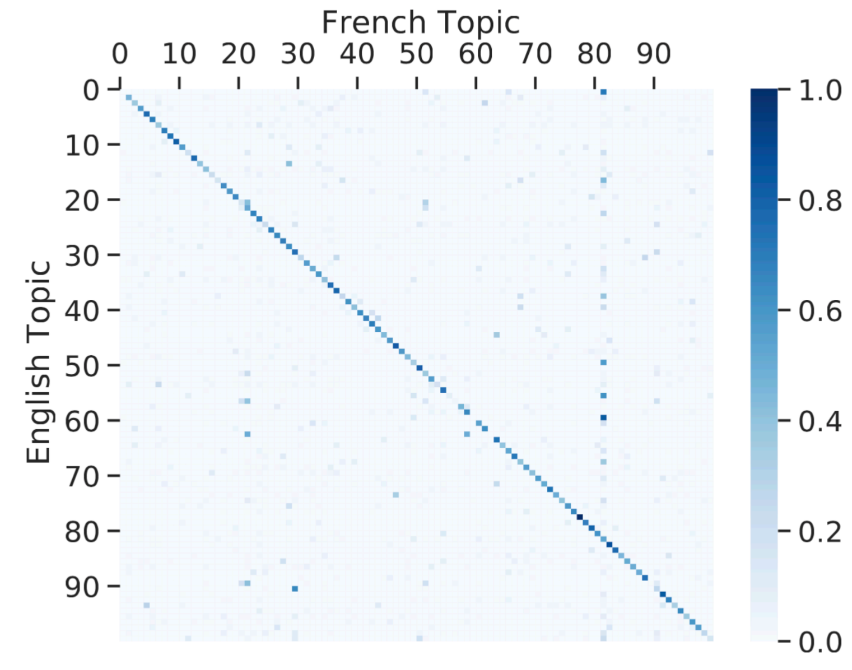
Turns out that embedding quality (as measured by Semantic Textual Similarity performance) is very important for cross-lingual transfer ($\rho = 0.93$) but not for topic coherence ($\rho = 0.46$).



What is fine-tuning doing to improve cross-lingual transfer?



Before fine-tuning: model selects foreign doc's topic from small subset of available English topics.



After fine-tuning: much better!

5. Takeaways

1. Fine-tuning embeddings is **essential** for contextualized neural topic modeling.
2. Best topics achieved through (TC)CTMs fine-tuned on or optimized over topic classification.
3. Best topic transfer achieved through CTMs fine-tuned on NLI.
   - ➡ Strong correlation between embedding quality and topic transfer (but it saturates quickly).
   - ➡ Embedding quality and topic quality not strongly correlated, so don't directly optimize over STS!
4. Continued pre-training is counterproductive for this task.

More analyses in the paper!