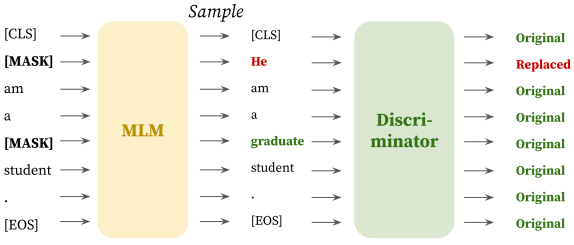
Prompting ELECTRA: Few-Shot Learning with Discriminative Pre-Trained Models

Mengzhou Xia, Mikel Artetxe, Jingfei Du, Danqi Chen, Ves Stoyanov

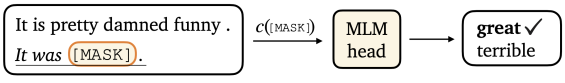
mengzhou@princeton.edu

Motivation

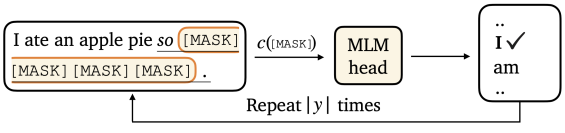
- Discriminative models are strong alternatives to masked language models as universal encoders



- Previous works propose prompt-based learning with MLMs (Gao et al., 2021, Schick and Schütze, 2021)



Single-token classification tasks

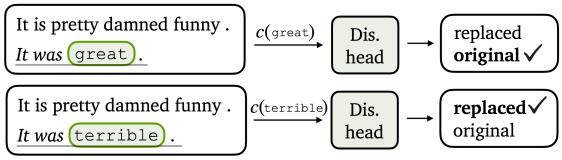


multi-token multiple-choice tasks

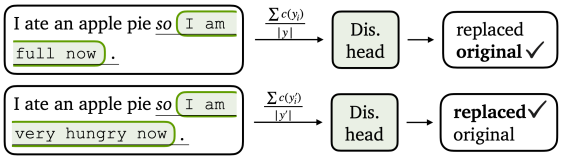
- How can we perform prompt-based learning with discriminative models?

Methodology

- Reuses the discriminative pre-trained objective
- If a token/option has a higher probability of being original, then it's correct.



Single-token classification tasks

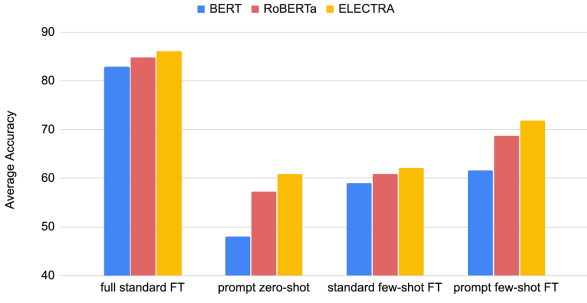


multi-token multiple-choice tasks

- prob: average probability of all tokens in an option
- rep: feed the average representation of all tokens in an option to the discriminator

Results

- 9 Single-token classification tasks
- ELECTRA outperforms BERT and RoBERTa by a large margin on zero-shot and few-shot prompt-based learning.

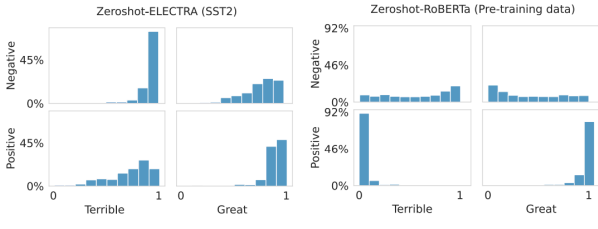


- 4 multiple-choice tasks
- ELECTRA outperforms RoBERTa.

	COPA	StoryCloze	HellaSwag	PIQA
RoBERTa-base				
[CLS]	69.3 (2.1)	81.4 (1.2)	32.1 (2.3)	54.6 (1.0)
PET	**78.1 (2.6)**	69.8 (3.9)	30.1 (2.1)	61.9 (1.2)
ELECTRA-base				
[CLS]	76.0 (0.8)	84.5 (1.1)	**58.8 (2.0)**	64.9 (1.1)
prob	76.7 (1.7)	85.7 (1.5)	54.5 (1.0)	65.8 (0.3)
rep	76.7 (1.7)	**86.6 (1.1)**	58.0 (0.7)	**66.2 (0.4)**
RoBERTa-large				
[CLS]	76.0 (1.4)	86.5 (4.9)	44.1 (5.1)	55.5 (1.6)
PET	85.4 (2.9)	85.4 (5.3)	46.9 (5.0)	64.6 (3.9)
ELECTRA-large				
[CLS]	**96.0 (0.8)**	**95.6 (0.7)**	**79.9 (2.6)**	71.2 (3.8)
prob	89.0 (1.6)	90.2 (0.5)	77.9 (0.6)	72.0 (0.7)
rep	88.7 (0.9)	90.4 (0.6)	78.1 (0.5)	**72.7 (1.3)**

Why is ELECTRA better at prompting?

- ELECTRA's generator has the same probability for great and terrible when the ground truth is terrible, feeding the effective negatives to the discriminator.



Codebase: https://github.com/facebookresearch/ELECTRA-Fewshot-Learning