cBERT: A Meta-Learning Approach to Contextual Representations for Continual Learning

William Huang   wh629

Anhthy Ngo   an3056

Christine Shen   zs1534

Abha Sahay   as13492

Continual Learning

Constantly learning without "catastrophically forgetting" past experiences
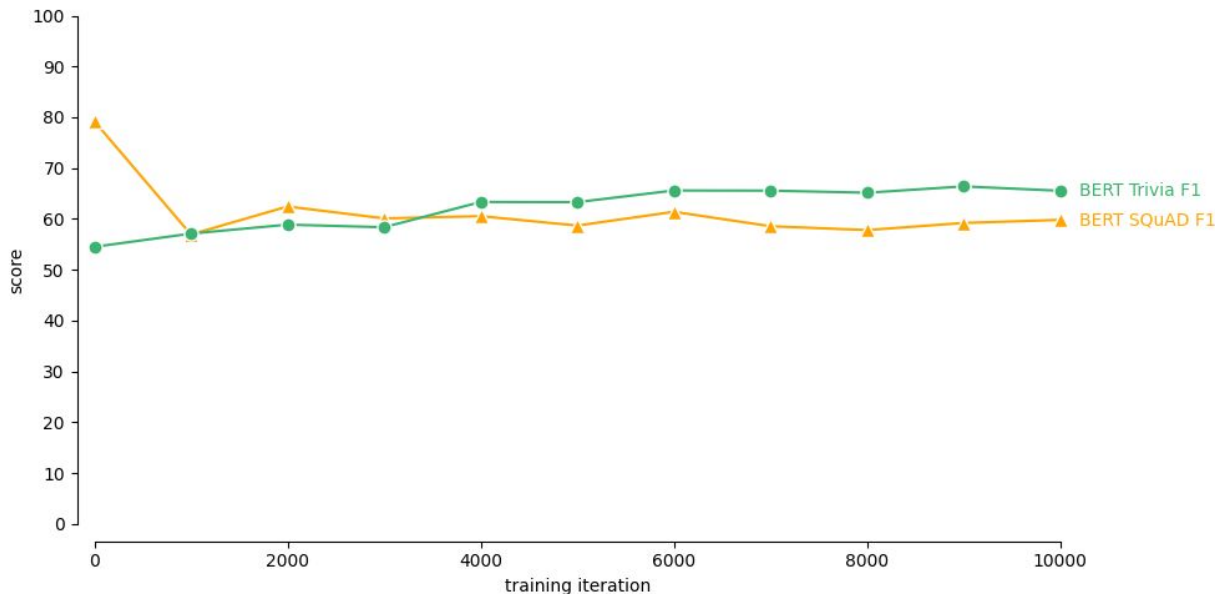
Estoy muy cansado → I am very tired

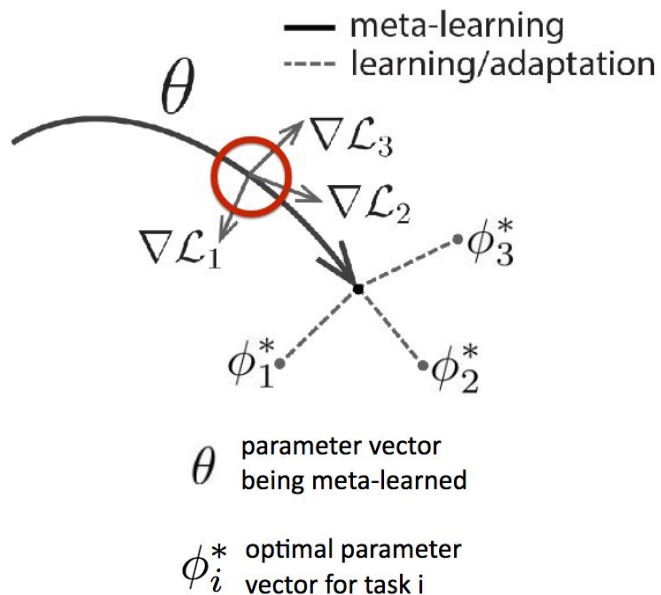Eu estou muito cansado → I am very tired

Estoy muy cansado → **?**

BERT for Continual Learning

- We fine-tune BERT on TriviaQA for 10,000 iterations and evaluate the fine-tuned embeddings on SQuAD every 1,000 iterations. As the model improves on TriviaQA, we see a dip in performance on SQuAD of about 18.9%.

Meta-Learning

— meta-learning
---- learning/adaptation

θ

$\nabla \mathcal{L}_3$

$\nabla \mathcal{L}_2$

$\nabla \mathcal{L}_1$

ϕ_3^*

ϕ_1^* ϕ_2^*

θ parameter vector being meta-learned

ϕ_i^* optimal parameter vector for task i

- **Idea:** Effectively use similar data to learn an inductive bias

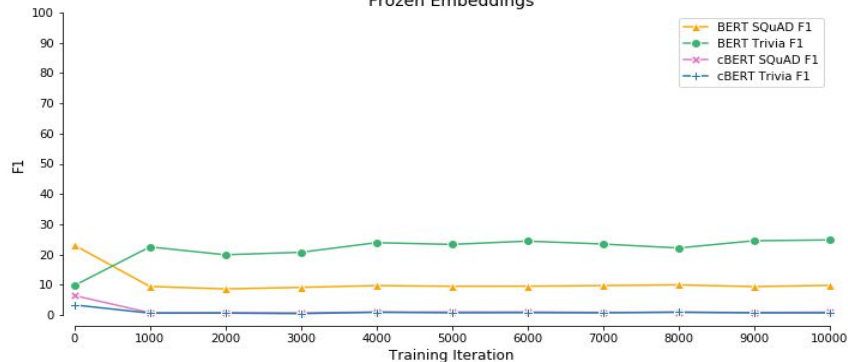# Can meta-learning help BERT with continual learning?

- Online aware Meta-Learning (OML)[1]

- Continual Learning Curriculum
  - SQuAD → TriviaQA

- Meta-Training
  - NewsQA, HotpotQA, and Natural Questions
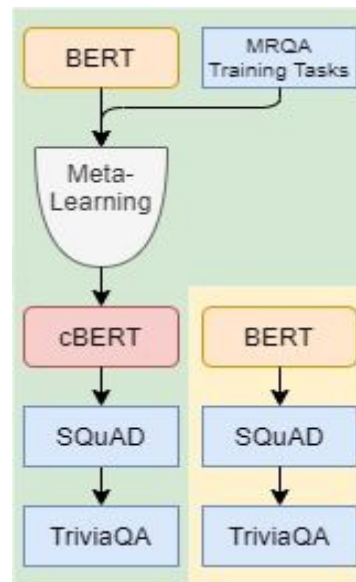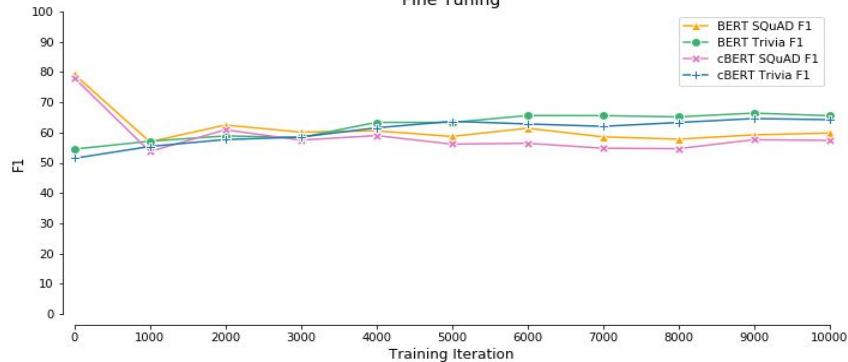

[1] Javed and White 2019

Findings

Frozen Embeddings

Embedding	Task	F1 (Max / Min)	Difference
BERT	SQuAD	23.1% / 8.7%	-14.5%
BERT	TriviaQA	24.9% / 9.8%	+15.1%
cBERT	SQuAD	6.5% / 0.8%	-5.7%
cBERT	TriviaQA	3.4% / 0.5%	-2.9%

Fine Tuning

Embedding	Task	F1 (Max / Min)	Difference
BERT	SQuAD	79.2% / 56.9%	-22.3%
BERT	TriviaQA	66.4% / 54.5%	+11.9%
cBERT	SQuAD	78.0% / 53.7%	-24.3%
cBERT	TriviaQA	64.6% / 51.5%	13.1%

Thank You