Héctor Javier
hjvm@mit.edu

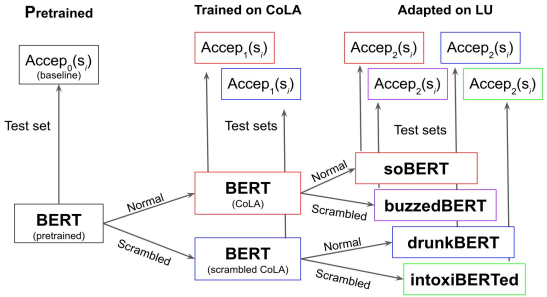
BERT'S ADAPTABILITY TO SMALL DATA

Annika Heuser
aheuser@mit.edu

VISION

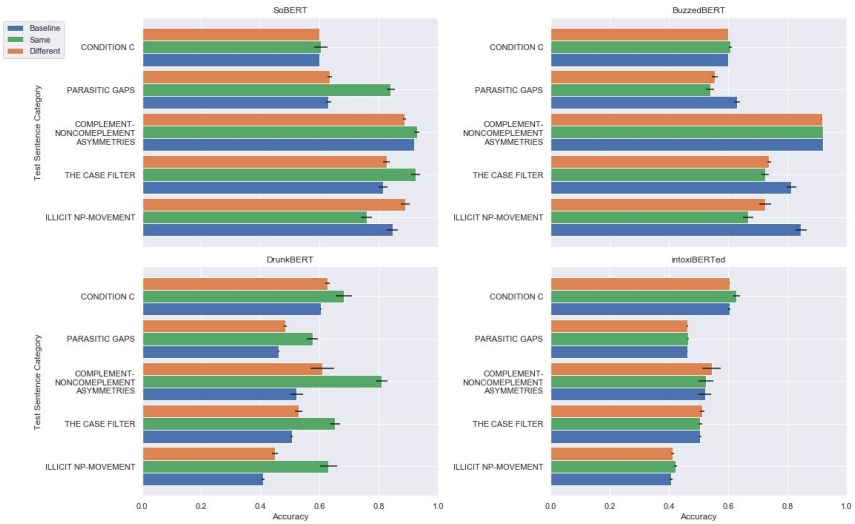
In order to characterize the stability of the apparent Knowledge of Language (KOL) BERT(Devlin et al., 2018) exhibits in downstream NLP tasks, we must understand how sensitive BERT's performance is to small regularities in the fine-tuning data.

STEPS



After observing the effect of training BERT on 8,551 scrambled or regular CoLA (Warstadt et al., 2018) sentences, we investigate the added effect of training on 10 regular or scrambled adaptation minimal pairs.

NEWS



The 20 adaptation sentences (Prasad et al., 2019) (Kodner and Gupta, 2020) give DrunkBERT a statistically significant ($p << 0.001$) jump in performance across all 5 syntactic phenomena.

4-gram models trained in the same way only saw modest performance improvements in comparison, suggesting that lexical cues are not responsible.

CONTRIBUTIONS & ONGOING WORK

- Proposed experimental paradigm to gauge the stability of BERT's KOL.
- Demonstrated BERT's tendency to reflect the regularities of the data used for the most recent round of fine-tuning
- Call for consideration of how observed KOL in BERT may be more a function of the data used for fine-tuning than the model itself.

Abstract:

Massachusetts Institute of Technology