Morphology Improves CLMs

- Character language models (CLMs) have the capacity to share subword information across morphological forms.
- Hypothesis: accurately modeling morphology improves CLM performance, but it is difficult for CLMs to learn this from the language modeling objective alone.
- We incorporate morphology annotations into a CLM using a multi-task objective.
- Adding morphology into CLMs improves bits-per-character (BPC) across 24 languages, even when the LM and morphology data is disjoint.

Effect of Training Data Quantity

Language Modeling Results

- Above: BPC performance on MWC test set of HCLM (best model from Kawakami et al., 2017), our baseline LM, and the MTL model.
- Right: Improvement in BPC of the MTL model over the LM baseline for 24 languages.
- Find that across all languages, BPC improves when morphology supervision is added to vanilla CLM.

What drives this improvement in CLM performance?

- Incorporate morphology supervision into model via multi-task learning:
  \[ L(e, m) = L_{CLM}(e) + \delta \sum_{i=1}^n L_{CLM}(m) \]
- The LM architectures consist of a stacked LSTM model with the layer at which we multi-task morphology selected as a hyperparameter.
- We train both baseline LMs (LM) and multitasked LMs (MTL) on the text of Universal Dependencies (UD) for 24 languages, as well as on the Multilingual Wikipedia Corpus (MWC).
- UD morphological features are used as supervision for all MTL models.

Cross-Lingual Transfer

- Can morphology from a high-resource language improve LMs on a low-resource, typologically similar language?
- We incorporate additional data from a high-resource language into CLMs for a related, low-resource one.
- Find that adding both LM data and morphological features helps model the low-resource language.