Advantages Brought by Scaling

- Low-rank parameterization enables a dramatic increase in the numbers of nonterminals (NT) preterminals, from just over 30 and 60 to upwards to 5,000 and 10,000 respectively.
- The Sentence-F1 score in unsupervised parsing sees an increase from 55.2 to 64.1, a significant improvement attributable to scaling.

The Achilles’ Heel of Low-rank PCFGs

- Despite benefiting from scaling in unsupervised parsing, low-rank PCFGs perform poorly as a language model and underperform similarly-sized HMMs.
- On the Penn Treebank, PCFGs scaled via low-rank parameterization with thousands of states achieve F1 = 170.
- However, it lags behind a similarly-sized HMM which obtains $\approx 130$ perplexity, even though HMMs are subclass of PCFGs.

SimplePCFG

- In simple PCFGs, we simplify all these things.
- In simplePCFG, we parameterize $L, R$ directly instead of through the shared $U^T$, which in fact contributes to building a more flexible parameterization.

Low-Rank PCFGs

- The previous approach to scaling HMMs and PCFGs to thousands of nonterminals is parameterizing the rule probability tensor $T \in \mathbb{R}^{[N] \times [S] \times [N]}$ to be low-rank.
- Low-rank PCFGs can be viewed as introducing a new latent variable, namely a “rank variable” $R$, where $U, V, W$ are tensor/matrix representations of rule probabilities.
- In fact, a low-rank PCFG can be parameterized as a PCFG with independent left/right productions by marginalizing nonterminal variables and viewing the rank variables as new nonterminal variables.
- As such, low-rank PCFGs parameterize $L, R$ in a more restrictive manner: $L = UVT, R = WUT$. We speculate that the shared $U^T$ would restrict the expressiveness of low-rank PCFGs and thus hinder optimization, which motivates our simple PCFGs.

Language Modeling

- Simple Neural PCFG (SN-PCFG) outperforms previous low-rank PCFG with similar size.
- SN-PCFG successfully exceeds similarly-sized HMMs.

Speed & Memory Comparison

- log-sum-exp: 512 20 1x 100x
- log-einsum-exp: 512 20 4.8x 3x
- FlashInside: 512 20 9.5x 1x
- log-einsum-exp (FlashInside): 8192 20 1x 2x
- log-einsum-exp (FlashInside): 8192 20 6x 1x
- log-einsum-exp (FlashInside): 512 40 1x 50x
- log-einsum-exp (FlashInside): 512 40 16x 3x
- FlashInside: 8192 40 44x 1x
- log-einsum-exp (FlashInside): 8192 40 1x 2.4x
- FlashInside: 8192 40 39x 1x

Unsupervised Parsing

- SC-PCFG is simple compound neural PCFG.
- SN-PCFG or SC-PCFG outperforms previous PCFG models on unsupervised parsing benchmarks across different languages.
- Simple PCFG vs. Neural PCFG: Despite the better scalability of simple PCFGs, we find that under the same number of NT (i.e., 128), SN-PCFG expectedly underperforms N-PCFG.