Finding Sparse Structure for Domain Specific Neural Machine Translation

Jianze Liang, 1,2∗ Chengqi Zhao, 2 Mingxuan Wang, 2 Xipeng Qiu, 1 Lei Li 2
1 Fudan University 2 ByteDance AI Lab

Prune-Tune: An Effective and Flexible Schema for Domain Adaptation in NMT

Effective Pruning for Transformer

Robust Training

Effective for Low-resource Domain Adaptation

Keeping General Knowledge to better Learn the Target Domain

Few parameters are needed to train most target domains

Sequential Multi-Domain Adaptation: Learning without Forgetting

Table 2: BLEU results of domain adaptation on En→De:

Pruning Rate | WMT | EWSLT | EMEA | Novel | #Pruning Para.
---|---|---|---|---|---
10% | 28.7 | 32.3 | 30.6 | 24.3 | 30.9 | 24.3
30% | 28.3 | 32.4 | 30.3 | 23.8 | 30.9 | 24.3
50% | 28.1 | 32.2 | 29.5 | 23.6 | 30.8 | 24.3
70% | 26.8 | 31.1 | 28.9 | 23.1 | 30.4 | 24.3

Table 3: BLEU Results of Sequential Domain Adaptation on En→De. (#M denotes the number of required models. W, I, E, N refer to datasets WMT14, EWSLT, EMEA, Novel, respectively. In our Sequential P-Tune Model, general domains occupied 30% parameters, and each target domain occupied 10%.)

Table 4: BLEU Results of Sequential Domain Adaptation on En→Zn. (#M denotes the number of required models. In our Sequential P-Tune Model, general domains occupied 30% parameters, and each target domain occupied 5%).