Handling Syntactic Divergence in Low-resource Neural Machine Translation

Supplementary Materials

A Results

We present the full results in Tab. 1, from which we can see that as the amount of supervised data increases, the performance gain of SMT is not as much as the NMT model. For SMT, reordering has much better performance than no-reorder, but still lags behind the supervised counterpart.

	3k		6k		10k		20k		400k		ug	
Model	NMT	SMT	NMT	SMT	NMT	SMT	NMT	SMT	NMT	SMT	NMT	SMT
sup	2.17	6.36	7.86	8.70	11.67	10.68	15.98	12.11	26.56	18.62	0.58	1.46
back	2.27	8.46	5.40	10.61	13.50	12.05	16.05	13.68	–	–	0.42	1.37
No-Reorder	6.46	3.08	9.73	5.24	12.57	6.72	15.56	8.96	–	–	3.24	1.67
Reorder	**9.94**	6.23	**12.42**	8.14	**14.98**	9.22	**17.58**	11.21	–	–	**4.17**	1.07

Table 1: BLEU of our approach (Reorder) with different amount of parallel sentences of `ja-en` and `ug-en` translation. Baselines are supervised learning (sup), supervised learning with back translation (back) and data augmentation with translated original English sentences (No-Reorder).