Neural System Combination for Machine Transaltion

Long Zhou, Wenpeng Hu, Jiajun Zhang and Chengqing Zong
{long.zhou, wenpeng.hu , jjzhang, cqzong}@nlpr.ia.ac.cn

Background and Motivation

Neural machine translation (NMT) generates much more fluent results compared to statistical machine translation (SMT). However, SMT is usually better than NMT in translation adequacy. System combination is therefore a promising direction to unify the advantages of both NMT and SMT.

Our solution: **Neural System Combination** (**NSC**)

- Step 1: for a source sentence, generate translations with phrase-based SMT (PBMT), hierarchical phrase-based SMT (HPMT) and NMT.
- Step 2: design a multi-source sequence-to-sequence model that takes as input the three translation results of PBMT, HPMT and NMT, and produces the final target language translation.

Translation Example:

Source:	海珊 也 与 恐怖 组织网 建立 了 联系 。
Pinyin:	*hanshan ye yu kongbu zuzhiwang jianli le lianxi 。*
Ref.:	hussein has also established ties with terrorist networks.
PBMT:	hussein also has established relations and terrorist group .
HPMT:	hussein also and terrorist group established relations .
NMT:	hussein also established relations with <UNK> .
NSC:	hussein also has established relations with the terrorist group .

Table 1: Translation examples of single system and our model.

Overview of our approach

Inputs: translation outputs of PBMT, HPMT and NMT.

Outputs: final target language translation results.

Core idea: hierarchical attention-based multi-source seq2seq model.

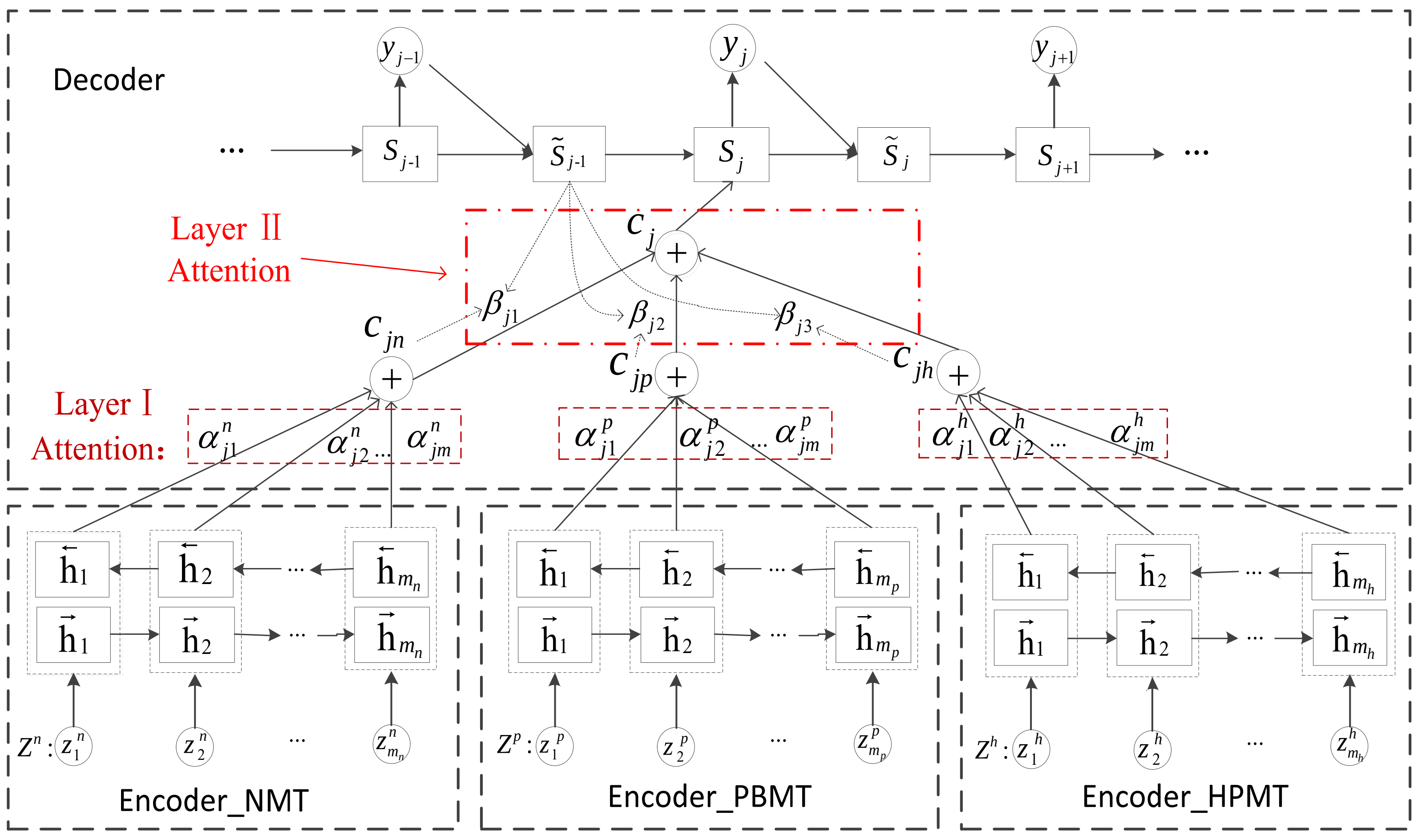
Decoder
y_{j-1} y_j y_{j+1}
s_{j-1} $\tilde{s}_{j-1}$ s_j $\tilde{s}_j$ s_{j+1}
Layer II Attention
c_j β_{j1} β_{j2} β_{j3}
c_{jn} c_{jp} c_{jh}
Layer I Attention:
α^n_{j1} α^n_{j2} α^n_{jm} α^p_{j1} α^p_{j2} α^p_{jm} α^h_{j1} α^h_{j2} α^h_{jm}
$\overleftarrow{h}_1$ $\overleftarrow{h}_2$ $\overrightarrow{h}_1$ $\overrightarrow{h}_2$
Z^n Z^p Z^h
Encoder_NMT
Encoder_PBMT
Encoder_HPMT

Figure 1: The architecture of neural system combination model

Layer I Attention:

$$c_{jk} = \sum_{i=1}^{m} \alpha_{ji}^{k} h_i \quad \Longleftarrow \quad \alpha_{ji}^{k} = \frac{\exp(e_{ji})}{\sum_{l=1}^{m} \exp(e_{jl})}$$

where $e_{ji} = v_a^T \tanh(W_a \tilde{s}_{j-1} + U_a h_i)$ scores how well $\tilde{s}_{j-1}$ and h_i match.

Layer II Attention:

$$c_j = \sum_{k=1}^{K} \beta_{jk} c_{jk} \quad \Longleftarrow \quad \beta_{jk} = \frac{\exp(\tilde{s}_{j-1} \cdot c_{jk})}{\sum_{k'} \exp(\tilde{s}_{j-1} \cdot c_{jk'})}$$

Training Data Simulation:

The neural system combination framework should be trained on the outputs of multiple translation systems and the gold target translations. In order to keep consistency in training and testing, we design a strategy to simulate the real scenario.

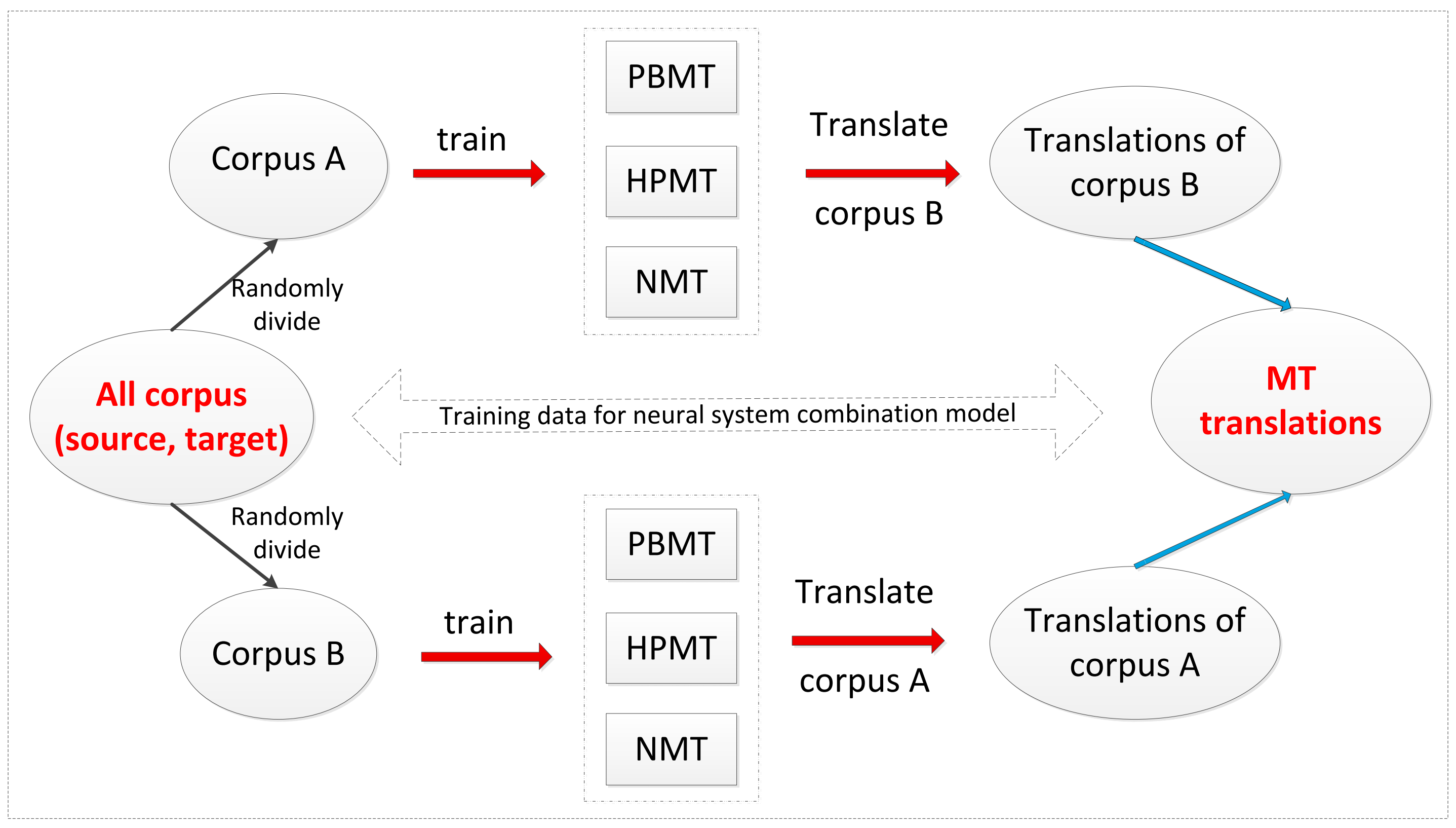

Figure 2: Strategy of training data simulation

Experiments

Data sets: 2.08M sentence pairs of Chinese-English extracted from LDC corpus.

Systems: we compare our neural combination system with the best individual engines, and the state-of-the-art traditional combination system Jane.

System	MT03	MT04	MT05	MT06	Ave
PBMT	37.47	41.20	36.41	36.03	37.78
HPMT	**38.05**	**41.47**	**36.86**	36.04	**38.10**
NMT	37.91	38.95	36.02	**36.65**	37.38
Jane	39.83	42.75	38.63	39.10	40.08
NSC	40.64	44.81	38.80	38.26	40.63
NSC+Source	42.16	45.51	40.28	39.02	41.75
NSC+Ensemble	41.67	45.95	40.37	39.02	41.75
NSC+Source+Ensemble	**43.55**	**47.09**	**42.02**	**41.10**	**43.44**

Table 2: Translation results (BLEU score) for different machine translation and system combination methods. Jane is an open source system combination toolkit that uses confusion network decoding.

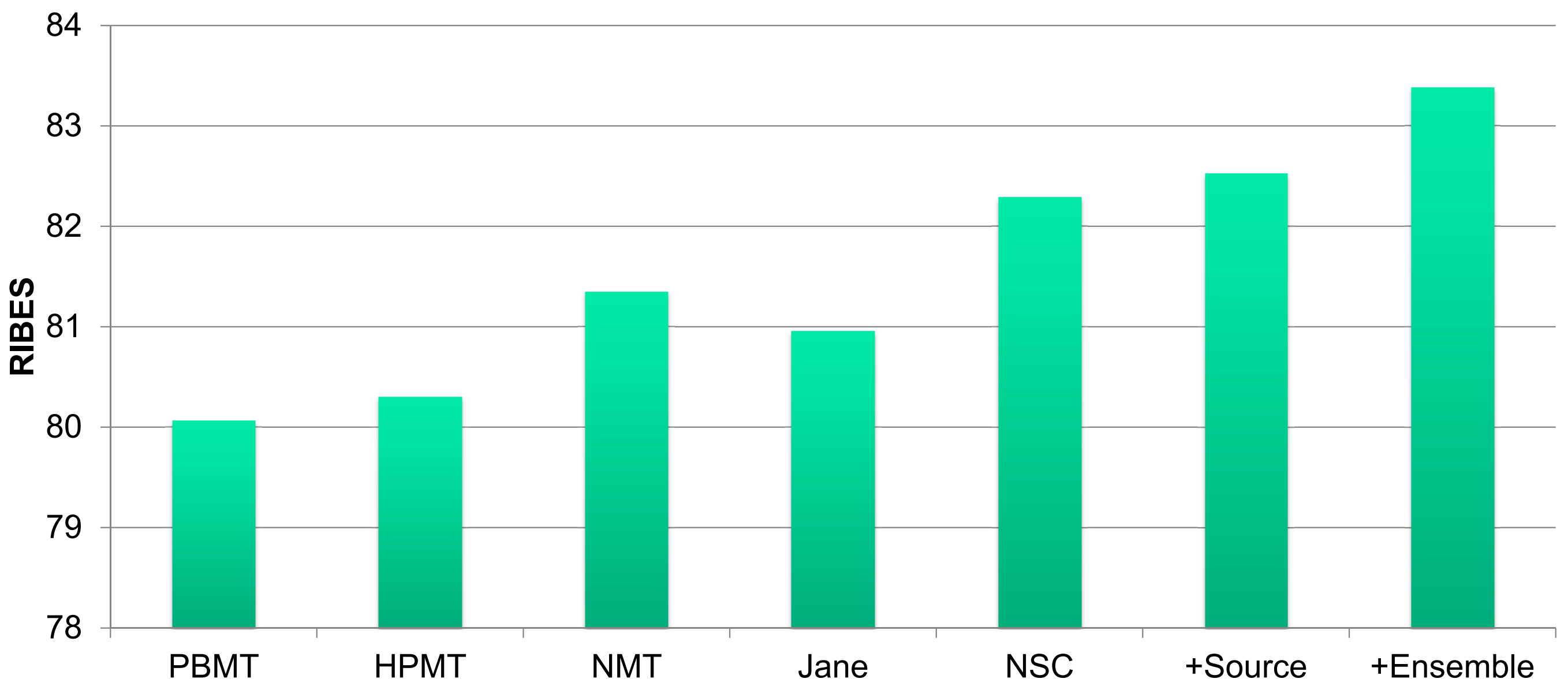

Figure 3: Comparison of translation fluency (word order), according to the automatic evaluation metrics RIBES.

Conclusions

- The proposed neural system combination method using hierarchical attentional seq2seq model can substantially improve the translation quality by combining the merits of SMT and NMT.
- Neural system combination architecture is simple and can be applied into other applications, such as summarization and text generation.

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences