


 ADAPT

 Engaging Content
 Engaging People

APE through neural and statistical MT with augmented data

ADAPT/DCU submission to the WMT 2019 APE Shared task

Dimitar Shterionov

Joachim Wagner

Félix do Carmo

 ADAPT Centre, School of Computing, Dublin City University, Dublin, Ireland
 [dimitar.shterionov, joachim.wagner, felix.docarmo]@adaptcentre.ie

Motivation

- Extra context (specific global properties of groups of segments) added as a prefix or a suffix to each segment
- Successful in domain adaptation of MT and APE, this technique deserves further attention

Prefixes per

 Meaning: **Topic models** Structure: **Sentence length**

- Interleaving different technologies for MT (e.g. neural) and APE (e.g. phrase-based statistical) can lead to better results

Approach(es)

- Neural:
 - multi-source systems trained on context-augmented
 - Marian-NMT (multi-s2s) with LSTM units
 - Early stopping after 5 epochs
- Statistical:
 - monolingual: original MT output → post-edited variant
 - Moses with Giza++ for word alignment
 - 5-gram KenLM language model, tuned with Mert

Bins (for NPE)

Topic models

- Latent Dirichlet Allocation (LDA)
- On the source (English) side
- Ten topic clusters

Sentence length

- # of tokens in the source sentence
- 8 bins of similar sizes

Input augmentation:

<TOPIC1> In addition, four-color graphics using different hues are included.

<TOPIC1> Darüber hinaus werden vier Grautöne mit unterschiedlichen Grautönen verwendet.

Data

- Authentic and synthetic data
- Divided into 3 datasets for EN-DE and 2 datasets for EN-RU

Size	EN-DE	EN-RU
small	268 840	301 780
medium	795 208	N/A
large	4 660 020	8 037 141

Table 1: Number of SRC-NMT-PE triplets distributed over three data sets.

Size	EN-DE			EN-RU		
	SRC	NMT	PE	SRC	NMT	PE
small	10 771	15 477	18 088	9 125	14 783	15 761
medium	48 227	48 257	48 869	N/A		
large	50 327	50 538	50 790	53 030	50 646	52 970

Table 2: Vocabulary sizes (after applying BPE).

Results and Analysis

	Model	Prefix	BLEU ↑	TER ↓
MT	Baseline	N/A	76.94	15.08
NPE	small	N/A	63.28	24.09
	medium	N/A	70.57	18.81
	large	N/A	70.29	19.89
	small	topic	60.41	28.59
	medium	topic	73.08	17.81
	large	topic	75.82	15.89
	small	length	62.56	26.91
	medium	length	73.74	17.26
	large	length	75.85	15.91
	SMT	small	N/A	76.82
medium		N/A	77.04	15.17
large		N/A	76.82	15.26

Table 3: BLEU and TER scores for the EN-DE NPE and SMT models (dev set). Rows in bold indicate submitted system results.

	Model	Prefix	BLEU ↑	TER ↓
MT	Baseline	N/A	80.22	13.13
NPE	small	N/A	50.76	34.45
	large	N/A	59.01	28.01
	small	topic	48.30	41.19
	large	topic	75.39	16.18
	small	length	44.68	44.57
	large	length	73.67	19.74
	SMT	small	N/A	79.40

Table 4: BLEU and TER scores for the EN-RU NPE and SMT models (dev set). Rows in bold indicate submitted system results.

* Results for the large-topic system were obtained after the submission.

- NPE: systems are unable to perform as well as the original NMT systems
- INPE: increasing amounts of training data lead to the best scores when combined with context prefix
- NPE: The addition of context tokens has significant positive effect
- NPE: Topic information enables highest scores for EN-RU. For EN-DE *topic* and *length* lead to almost the same scores
- APE systems using SMT score highest, beating all neural ones.
- SMT: scores are close to the original ones, and consistent
- SMT: not sensitive to the size of the training data