The University of Maryland's Kazakh-English Neural Machine Translation System at WMT2019

Eleftheria Briakou and Marine Carpuat

University of Maryland

INTRODUCTION

✦ **Shared Task**

- ✦ News translation task
- ✦ Kazakh → English
- ✦ Low-resource MT

✦ **Submission Overview**

- ✦ Joint-BPEs [1]
- ✦ Transfer learning [2]
- ✦ Back-Translation [3]
- ✦ Ensemble

APPROACH

✦ Baseline model: NMT trained only on Kazakh-English parallel data

✦ Train on additional data

- **Parallel from Transfer Learning**
  - Parent language: **Turkish** → low-resource but related language
  - Child language: **Kazakh**
- **Synthetic from Back-translation**

Encoding	Original		Romanized	
Word	molekül	молекула	molekuel	molekula
BPEs	mol_ek_ül	мол_ек_ ул_а	mol_ek_uel	mol_ek_ula
Word	fosfor	фосфор	fosfor	fosfor
BPEs	fos_for	ф_ос_фор	fos_for	fos_for

lexical relatedness; different scripts

EXPERIMENTAL STUDY

① **different lexical representations;** maximize parameter sharing across languages

- ✦ Joint-Byte Pair Encoding learned on concatenation of child-parent(JBPEs) [4]
- ✦ Separate-Byte Pair Encoding learned separately on child-parent (SBPEs) [4]
- ✦ N-gam Encoding (Soft-Decoupled Encoding and variants) [5]

② **romanization**; increase overlap between Turkish and Kazakh

③ **synthetic training** data obtained by back-translation

EXPERIMENTAL SETUP

Neural Model

- Seq2Seq Encoder-Decoder
- 1-layer LSTM with attention

Data

Parallel:

- Kazakh-English: ~100K from News Commentary Corpus and English-Kazakh crawled corpus
- Turkish-English: ~200K from Setimes2 Corpus

Monolingual:

- English: ~100K from News Commentary corpus

Preprocessing

- Tokenization and Truecasing
- 32K merge operations for BPEs
- 64K 5-grams for N-gram encoding

RESULTS

Method	Romanization	Overlap
JBPEs	✓	0.44
	✗	0.13
SBPEs	✓	0.18
	✗	0.04
N-gram	✓	0.61

Method	Original	Romanized
Baseline	4.33 ± 0.16	4.49 ± 0.02
JBPEs	$\mathbf{9.35 \pm 0.10}$	$\mathbf{9.89 \pm 0.14}$
SBPEs	7.10 ± 0.26	9.70 ± 0.28
N-gram	–	9.17 ± 0.21

Method	Synthetic	BLEU
Transfer		**9.89**
+ Back-Translation	✓	9.38
+ ensemble(4)	✓	**9.94**

Submitted System

CONCLUSIONS

① transfer learning benefits BLEU even when transferring from **low-resource and related** language

② best BLUE → **maximum surface-level parameter sharing** between child-parent language pairs

③ NMT **sensitive to the quality of the back-translation**

REFERENCES

[1] Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. In ACL, 2016.

[2] Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. 2016. Transfer learning for low-resource neural machine translation. In EMNLP, 2016.

[3] Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Improving neural machine translation models with monolingual data. In ACL.

[4] Graham Neubig and Junjie Hu. 2018. Rapid adaptation of neural machine translation to new languages. In EMNLP, 2018.

[5] Xinyi Wang, Hieu Pham, Philip Arthur, and Graham Neubig. 2019. Multilingual neural machine translation with soft decoupled encoding . In ICLR, 2019.