THE UNIVERSITY *of* EDINBURGH

Morphology in Neural Machine Translation

Rico Sennrich

Institute for Language, Cognition and Computation
University of Edinburgh

March 15 2016

Why Worry about Morphology in NMT?

What we want

- represent open word vocabulary with closed symbol vocabulary
- good generalization:
  share statistical strength shared between related word forms
- efficient computation

Byte pair encoding for word segmentation [Sennrich et al., 2016]

bottom-up character merging

- iteratively replace most frequent pair of symbols ('A','B') with 'AB'
- apply on dictionary, not on full text (for efficiency)
- output vocabulary: character vocabulary + one symbol per merge

word	freq
'l o w </w>'	5
'l o w e r </w>'	2
'n e w e s t </w>'	6
'w i d e s t </w>'	3

freq	symbol pair	new symbol

Byte pair encoding for word segmentation [Sennrich et al., 2016]

bottom-up character merging

- iteratively replace most frequent pair of symbols ('A','B') with 'AB'
- apply on dictionary, not on full text (for efficiency)
- output vocabulary: character vocabulary + one symbol per merge

word	freq
'l o w </w>'	5
'l o w e r </w>'	2
'n e w es t </w>'	6
'w i d es t </w>'	3

freq	symbol pair		new symbol
9	('e', 's')	$\rightarrow$	'es'

Byte pair encoding for word segmentation [Sennrich et al., 2016]

bottom-up character merging

- iteratively replace most frequent pair of symbols ('A','B') with 'AB'
- apply on dictionary, not on full text (for efficiency)
- output vocabulary: character vocabulary + one symbol per merge

word	freq
'l o w </w>'	5
'l o w e r </w>'	2
'n e w est </w>'	6
'w i d est </w>'	3

freq	symbol pair		new symbol
9	('e', 's')	$\rightarrow$	'es'
9	('es', 't')	$\rightarrow$	'est'

Byte pair encoding for word segmentation [Sennrich et al., 2016]

bottom-up character merging

- iteratively replace most frequent pair of symbols ('A','B') with 'AB'
- apply on dictionary, not on full text (for efficiency)
- output vocabulary: character vocabulary + one symbol per merge

word	freq
'l o w </w>'	5
'l o w e r </w>'	2
'n e w est</w>'	6
'w i d est</w>'	3

freq	symbol pair		new symbol
9	('e', 's')	$\rightarrow$	'es'
9	('es', 't')	$\rightarrow$	'est'
9	('est', '</w>')	$\rightarrow$	'est</w>'

Byte pair encoding for word segmentation [Sennrich et al., 2016]

bottom-up character merging

- iteratively replace most frequent pair of symbols ('A','B') with 'AB'
- apply on dictionary, not on full text (for efficiency)
- output vocabulary: character vocabulary + one symbol per merge

word	freq
'lo w </w>'	5
'lo w e r </w>'	2
'n e w est</w>'	6
'w i d est</w>'	3

freq	symbol pair		new symbol
9	('e', 's')	→	'es'
9	('es', 't')	→	'est'
9	('est', '</w>')	→	'est</w>'
7	('l', 'o')	→	'lo'

Byte pair encoding for word segmentation [Sennrich et al., 2016]

bottom-up character merging

- iteratively replace most frequent pair of symbols ('A','B') with 'AB'
- apply on dictionary, not on full text (for efficiency)
- output vocabulary: character vocabulary + one symbol per merge

word	freq
'low </w>'	5
'low e r </w>'	2
'n e w est</w>'	6
'w i d est</w>'	3

freq	symbol pair		new symbol
9	('e', 's')	→	'es'
9	('es', 't')	→	'est'
9	('est', '</w>')	→	'est</w>'
7	('l', 'o')	→	'lo'
7	('lo', 'w')	→	'low'
...			

BPE and Morphology

- subword boundaries are not morpheme boundaries
- still, model is often able to produce correct forms
- learning BPE on concatenation of (transliterated) source and target text improves consistency

system	sentence
source	health research institutes
reference	Gesundheitsforschungsinstitute
word-level (with back-off)	Forschungsinstitute
joint BPE	Gesundheits|forsch|ungsin|stitute
source	rakfisk
reference	ракфиска (rakfiska)
word-level (with back-off)	rakfisk → UNK → rakfisk
BPE	rak|f|isk → пра|ф|иск (pra|f|isk)
joint BPE	rak|f|isk → рак|ф|иска (rak|f|iska)

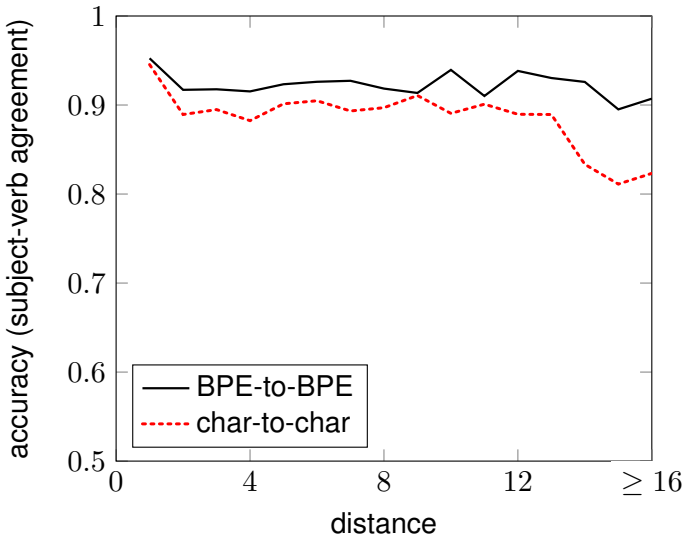
Assessing MT Quality with Contrastive Translation Pairs

Method [Sennrich, 2017]

compare probability of human reference translation with contrastive translation that introduces a specific type of error

Results

- character-level system [Lee et al., 2016] better than BPE-to-BPE system at transliteration, but worse at morphosyntactic agreement
- difference higher for agreement over long distances



Bibliography I

Lee, J., Cho, K., and Hofmann, T. (2016).

Fully Character-Level Neural Machine Translation without Explicit Segmentation.
ArXiv e-prints.

Sennrich, R. (2017).

How Grammatical is Character-level Neural Machine Translation? Assessing MT Quality with Contrastive Translation Pairs.
In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL), Valencia, Spain.

Sennrich, R., Haddow, B., and Birch, A. (2016).

Neural Machine Translation of Rare Words with Subword Units.
In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1715–1725, Berlin, Germany. Association for Computational Linguistics.