Segmentation Approaches and Machine Translation

Wei-Rui Chen

@UBC Computational Morphology RG

Outline

- Target-side Word Segmentation Strategies for Neural Machine Translation, Huck et al 2017
- Morphological Zero-Shot Neural Machine Translation, Zhou 2018

Target-side Word Segmentation Strategies for Neural Machine Translation

Matthias Huck, Simon Riess, Alexander Fraser

Huck et al 2017

- English -> German task
- Only German text is handled differently
  - Plain BPE for English text
- Plain BPE vs. morpho-preprocessing + BPE
- Morpho-preprocessing + BPE: 0.5 BLEU ⬆

Motivations – smaller |V|

- Vocabulary (V): a set of tokens generated (trained) by training text of source and target language either jointly or separately
- Smaller vocabulary size indicates
  - Lower memory requirement

Motivation – Data Sparsity

- Data Sparsity: A condition when large number of tokens in test text are not included in the vocabulary
- Avoid Data sparsity
  - Compared to word-level tokenization which will encounter out-of-vocabulary issue especially for a language that is morphologically complex, e.g. rich in compound words (German), sentence-word phenomenon (Inuktitut)

Motivation – Opener Vocabulary

- Open-vocabulary MT: the model is able to produce unseen words, e.g. new compounds, new stem+suffix combination

- There is ‘acceptable’, ‘adjustable’ in training data and the model learns to produce ‘enjoyable’ despite this is not in training data

Morpho-preprocessing

- Cascaded Application of Segmenters
- in -> suffix -> prefix -> compound -> out

- reviewing -> review+ing -> re+view+ing
- rebrainwashing -> rebrainwash+ing -> re+brainwash+ing -> re+brain+wash+ing
- extraordinary -> extraordin+ary -> ex+traordin+ary (if ‘ex’ is considered a prefix)

Morphological Processing + BPE

- How performing morpho-preprocessing can help BPE results
  - affixes can be learned separated from the stem
  - ex: reviewing -> rev + ie + wing by plain BPE if no morpho-preprocessing
  - Reviewing -> re + view + ing

Results

System	test2007		test2008	
	BLEU	TER	BLEU	TER
top 50K voc. (source & target)	25.5	60.9	25.2	60.9
BPE	25.8	60.7	25.6	60.9
compound + BPE	25.9	60.3	25.5	60.6
suffix + BPE	**26.3**	60.0	**26.0**	**60.1**
suffix + compound + BPE	26.2	**59.8**	25.8	60.2
suffix + prefix + compound + BPE	26.1	**59.8**	25.9	60.6
suffix + prefix + compound, 50K	25.9	59.9	25.5	60.3
phrase-based (Huck et al., 2015)	22.6	–	22.1	–

Morphological Zero-Shot Neural Machine Translation

Giulio Zhou

Morphological Zero-Shot Neural Machine Translation

- Challenges
  - ‘ir’ in ‘irregular’ is a prefix
  - ‘ir’ in ‘irrigate’ is not a prefix

  - ‘un’ in ‘unhappy’ is a prefix (adj)
  - ‘un’ in ‘union’ is not a prefix (N)
  - ‘un’ in ‘unhappiness’ is a prefix

Core idea

- They found Oversegmentation issue in prefix+suffix morphological segmenter (abbrv. Morph)

- Use mergeBPE to merge subwords back
  - union -> Morph -> un + ion
  - laboratory -> Morph -> lab + or + at + or + y
  - beautifully -> Morph -> beaut + i + ful + ly -> mergeBPE ->beaut + i + fully

Results

- BPE performs better in average BLEU

Segmentation		BPE	M.ssor	Morph	4k	16K	64k	128k
Multi	*tst2010*	**24.24**	23.49	23.17	23.53	23.85	24.08	24.10
	tst2017	**21.88**	20.64	20.36	21.04	21.05	21.37	21.45
Zero	*tst2010*	**23.66**	22.95	22.06	23.13	23.47	23.26	23.22
	tst2017	**17.89**	17.18	16.83	17.49	17.52	17.54	17.37

Table 4.3: Average BLEU scores on test sets for multilingual and Zero-Shot tasks with different segmentation strategies. Morph+MergeBPE is abbreviated with the number of merge operations.

Results

%	Model	Average BLEU		Average chrF3	
		tst2010	*tst2017*	*tst2010*	*tst2017*
10	Morph	13.60	10.66	35.62	32.49
	BPE	**13.77**	**10.8**	**35.67**	**32.54**
30	Morph	18.8	15.65	42.13	39.16
	BPE	**19.26**	**16.18**	**42.39**	**39.67**
50	Morph	20.43	17.72	44.39	41.45
	BPE	**21.14**	**18.03**	**44.44**	**41.65**
70	Morph	21.95	18.88	45.46	42.77
	BPE	**22.55**	**19.48**	**46.06**	**43.36**

Table 4.6: Average scores of all translation directions for low-resource (Zero-Shot) models (BPE vs Morph+64K)

Results

%	Model	de→nl		nl→de		it→ro		ro→it	
		tst2010	*tst2017*	*tst2010*	*tst2017*	*tst2010*	*tst2017*	*tst2010*	*tst2017*
10	Morph	12.07	8.62	11.74	7.73	8.53	6.86	9.17	8.05
	BPE	**12.4**	**8.63**	**12.45**	**8.56**	**8.61**	**6.87**	**9.57**	**8.28**
30	Morph	15.51	12.12	16.11	11.96	12.2	**11.09**	**13.10**	11.74
	BPE	**16.93**	**13.29**	**17.08**	**13.18**	**12.65**	10.91	12.74	**11.75**
50	Morph	**17.67**	**13.95**	**17.97**	13.61	**13.50**	**12.38**	14.58	13.31
	BPE	16.82	12.94	17.84	**14.12**	12.89	12.31	**15.07**	**14.95**
70	Morph	18.84	**15.40**	19.48	14.95	13.48	13.29	14.00	13.25
	BPE	**18.95**	14.93	**19.57**	**15.52**	**16.19**	**13.79**	**16.75**	**15.46**

Table 4.7: BLEU scores on Zero-Shot directions with different resource levels (BPE vs Morph+64K)

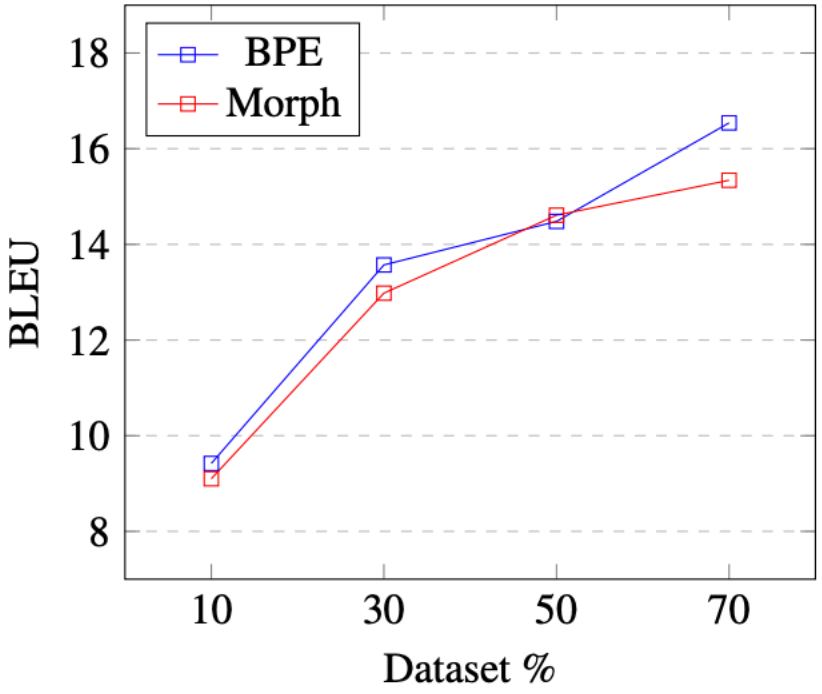

Figure 4.1: Average BLEU scores on Zero-Shot directions (both *tst2010* and *tst2017*) for BPE and Morph+64K

Discussion

- Vocabulary size and embedding representation
- Oversegmentation
- Why rule-based method not working

Discussion

- WordPiece
- In linguistics , morphology is the study of words , how they are formed , and their relationship to other words in the same language .
- SentencePiece
- _In _linguis tics , _ morph ology _is _the _study _of _words , _how _they _are _for med , _and _their _relationship _to _other _words _in _the _same _language .
- Porter Stemmer
- In linguist , morpholog is the studi of word , how they are form , and their relationship to other word in the same languag .