The JHU Machine Translation Systems for WMT 2017

Shuoyang Ding, Huda Khayrallah, Philipp Koehn, Matt Post, Gaurav Kumar, Kevin Duh

Center for Speech and Language Processing, Department of Computer Science

Phrase-Based Systems

- Best practices on all available data
- Och clusters for language model
- Operation sequence model
- Hierarchical reordering model
- Sparse lexical features
- No special handling of morphology (except: German)
- Huge language models trained on CommonCrawl data

Language	Tokens	LM Size
Czech	6.7 billion	13GB
German	65.2 billion	107GB
English	65.1 billion	89GB
Finnish	2.9 billion	8GB
Russian	23.3 billion	41GB
Turkish	11.9 billion	23GB

- Results

Language Pair	JHU 2016	Baseline	Och LM	Och+CC LM	Och+CC LM	Best NMT
	newstest2016				newstest2017	
English-Turkish	9.22	9.22	9.11	9.30	9.8	18.1 +8.3
Turkish-English	12.94	13.03	12.92	12.83	12.6	20.1 +7.5
English-Finnish	13.76	14.12	14.04	13.99	14.5	20.7 +6.2
Finnish-English	19.08	19.72	19.36	19.16	20.5	-
English-Latvian	-	18.66	18.71	18.85	14.4	20.1 +5.7
Latvian-English	-	25.82	26.03	26.12	16.8	20.0 +3.2
English-Russian	23.99	21.45		23.16	25.3	29.8 +4.5
Russian-English	27.88	24.47		27.22	31.5	34.7 +3.2
English-Czech	23.56			23.05	19.1	22.8 +3.7
Czech-English	30.37	29.84	29.98	29.80	26.5	30.9 +4.4
English-German	28.35	28.95		28.39	21.6	28.3 +6.7
German-English	34.50		34.20	33.87	29.7	35.1 +5.4

- Conclusion
  - vastly inferior to best NMT (by 3.2-8.3 BLEU points)
  - offering unconditional surrender

Rescoring with Neural Translation Model

- Rescore the output of Phrase Based MT (PBMT) with Neural MT (NMT)
- 500-best list rescoring (with Nematus)
- Lattice search: Khayrallah et al. (IJCNLP 2017)
  - Convert Moses search graph to OpenFST format
  - Remove epsilon arcs, determinize, minimize and topsort
  - Prune search graph
  - Stack decoding algorithm for lattice search using NMT model

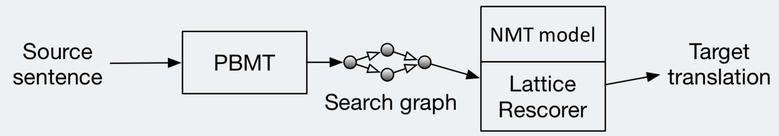

Figure: The neural lattice rescorer pipeline.

Language Pair	PBMT	NMT	NMT-Ens	N-best	Lattice	N-best	Lattice
	newstest2016					newstest2017	
English-Turkish	9.2	8.1	8.5	9.4	**9.9**	9.4	**10.4**
English-Finnish	14.1	12.6	13.6	14.6	**15.5**	14.3	**16.0**

Chinese-English Systems

- Built syntax-based and neural machine translation (NMT) systems
- Chinese tokenized with Stanford Chinese Word Segmenter
- Syntax-based (string-to-tree) system built with Moses decoder
- NMT systems built with Nematus
  - BPE on English and **tokenized** Chinese
  - 4-best Ensemble with model averaging
  - Continued training with back-translated 2 million samples for both English (News Crawl 2016) and Chinese monolingual data (XMU Xinhua News Corpus)

Language	Syntax	Base single	Base ensemble	Back-trans single	Back-trans ensemble
ZH-EN	16.22	17.81	**18.46**	17.52	18.16
EN-ZH	14.43	17.22	17.95	17.76	**18.60**