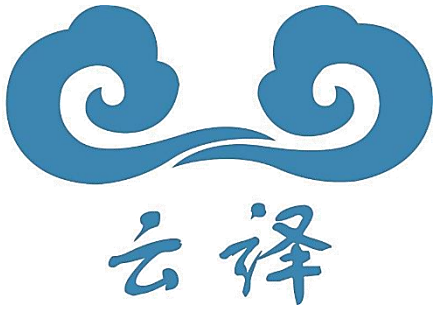

XMU Neural Machine Translation Systems for WAT 2017

Boli Wang, Zhixing Tan, Jinming Hu, Yidong Chen and Xiaodong Shi
Xiamen University & Yunyi Tech

XMU at WAT 2017

- Tasks
  - JIJI
  - IITB
  - Cookpad
- Models
  - NMTs for all tasks
  - DL4MT (our reimplementation)
  - open-sourced: XMUNLP@Github
- Results
  - Our NMTs outperform SMTs even in low-resourced tasks

NMT Systems

- NMT model
  - DL4MT (Our open-sourced reimplementation)

- Training
  - Optimizer: Adam
  - Early-stopping: initial α = 5e-4
  - Clip norm of gradients to 1.0
  - Dropout: keep rate = 0.8

- Decoding
  - Beam search (beam = 10)
  - Ensembles of 4 models
    - Different data shuffling and parameter initialization

Data Processing

- Tokenization
  - English
    - Moses tokenizer
    - Moses truecaser
  - Japanese
    - Convert full-width variants to half-width ASCIIs
    - mecab segmenter
  - Hindi
    - pre-tokenized by IITB

Data Processing

- Data filtering
  - By language
    - Unicode: Devanagari characters
    - langid toolkit
  - By word alignment score
    - fast-align toolkit

Data Processing

- Subword Segmentation
  - BPE
    - English & Hindi
    - Add suffix "@@" to non-final subword units
  - Mixed word/character model
    - Japanese
    - No suffix or prefix to subword units
  - 20K subwords for JIJI & IITB
  - 10K subwords for Cookpad

Data Processing

- Synthetic Training Data
  - IITB's monolingual corpus
  - Data selection by perplexity
    - Use srilm KN-5
    - Rank sentences by perplexity and choose the middle ones
    - 2.5M English sentences
    - 2.5M Hindi sentences
  - Back-translation
  - Mix synthetic bilingual data with sampled real bilingual data (about 1 : 1)

Results

- JIJI
  - Only 200K bilingual sentences
  - Single NMT outperforms SMT
  - Ensemble of 4 models improves +2.2 BLEU

System	EN-JA		JA-EN	
	BLEU	Human	BLEU	Human
HPBMT	16.22	10.25	15.67	10.25
Baseline	17.92	– –	15.77	– –
+Ensemble	**20.14**	**11.75**	**17.95**	**20.75**

Results

- IITB
  - 1.5M bilingual sentences + 2.5M monolingual sentences
  - Single NMT outperforms SMT (+2.9 BLEU)
  - Synthetic data really works (+ >6.0 BLEU)
  - Ensemble of 4 models improves +1.6 BLEU

System	EN-HI		HI-EN	
	BLEU	Human	BLEU	Human
PBMT	10.79	– –	10.32	– –
Baseline	13.69	– –	13.30	– –
+Synthetic	19.79	– –	20.61	– –
+Ensemble	**21.39**	**64.50**	**22.44**	**68.25**

Results

- Cookpad
  - 300K parallel pairs from six fields
    - title, step, ingredient
    - history, advice, and description
  - titles and steps
    - NMT >> SMT
    - where fluency matters
  - Ingredients
    - NMT > SMT
    - where adequacy matters

System	EN-JA BLEU	EN-JA Human	JA-EN BLEU	JA-EN Human
all				
PBMT	19.10	– –	23.87	– –
Baseline	22.47	– –	27.04	– –
+Ensemble	**24.44**	– –	**28.83**	– –
title				
PBMT	16.57	– –	9.72	– –
Baseline	16.90	– –	14.25	– –
+Ensemble	**18.78**	**23.75**	**15.57**	**10.25**
step				
PBMT	18.53	– –	22.84	– –
Baseline	22.01	– –	26.31	– –
+Ensemble	**24.00**	**45.50**	**28.03**	**40.50**
ingredient				
PBMT	29.60	– –	44.42	– –
Baseline	30.90	– –	43.89	– –
+Ensemble	**33.19**	**-3.75**	**46.98**	**3.50**

Thank you!