QCRI's Machine Translation Systems for IWSLT'16

Nadir Durrani Fahim Dalvi Hassan Sajjad Stephan Vogel

Qatar Computing Research Institute, HBKU

MOTIVATION

To build the best competition grade Arabic→English and English→Arabic machine translation systems.

DATA PREPARATION

Arabic being a morphologically rich language, requires some preprocessing for machine translation systems to perform well. Apart from tokenization on both languages, we also segment and normalize the Arabic data to reduce the sparsity of the source language.

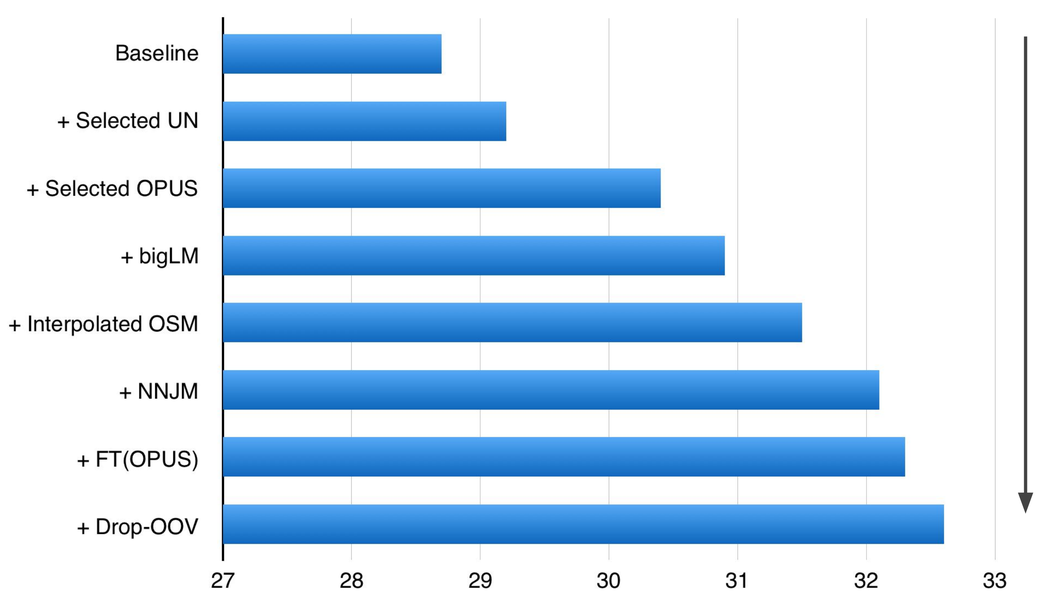
PHRASE BASED MACHINE TRANSLATION

System Settings

Moses Decoder | Max. sentence length: 80 | 5-gram OSM | 5-gram KENLM | Lex Reordering | 14-gram NNJM | Max. phrase length: 5 | K-best batch MIRA tuning

Key Experiments

Data Selection | Bigger Language Model | OSM Interpolation | NNJM Adaptation | Class-based Models | Drop OOV | Transliteration



Progress for Arabic→English Phrase based system
(cased BLEU)

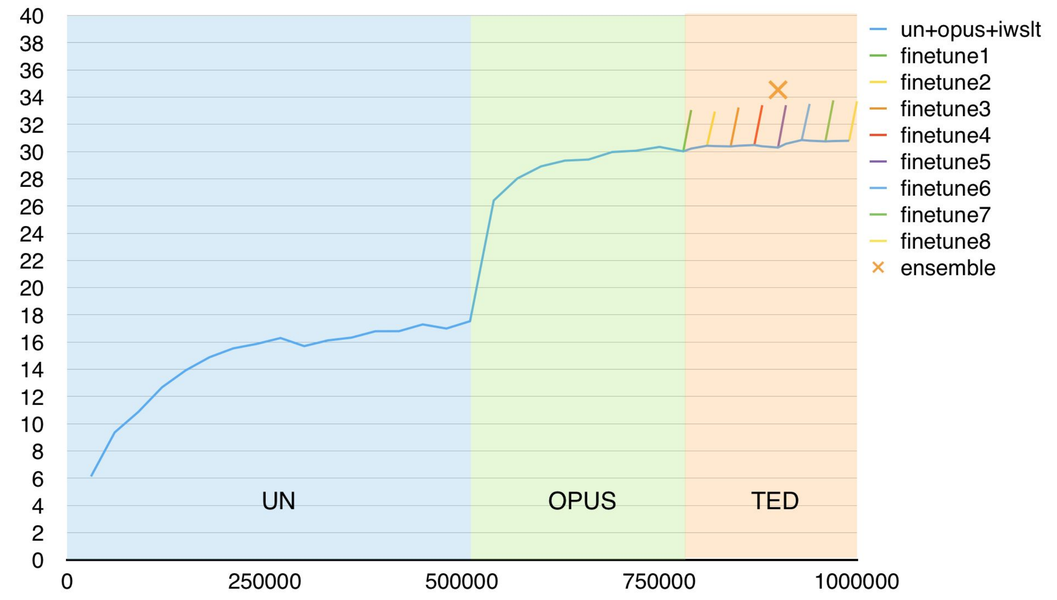
NEURAL MACHINE TRANSLATION

System Settings

Nematus | Max. sentence length: 80 | 59500 BPE Ops | 1024 LSTM units | Embedding size: 500 | Batch size: 80 | Bidirectional encoder | Attention-based model

Key Experiments

Data Selection | Finetuning Variants | Dropout | BPE model data | Ensemble | Data Concatenation



Progress for Arabic→English Neural system
(cased BLEU vs training iterations)

NEURAL MACHINE TRANSLATION KEY TAKEAWAYS

Data Selection

- Helps in Phrase based
- Hurts in Neural MT
  - Systems take much longer to train on full data, but overall performance is better

Finetuning

- Concatenation of all data hurts
- Training out-of-domain model first and then finetuning on in-domain works best

cat(in,out) < cat(in,out)→ft(in) < out→ft(in)

Layer Freezing and Dropout

- Freezing part of the network does not help
- Dropout only helps when applied to training on in-domain data

Advantages

- Adaptation is easier
- Final model size is independent of training data size
- Total training time is comparable, but human effort is greater for Phrase based

RESULTS ON DEVELOPMENT AND OFFICIAL SETS

System	ted-11	ted-12	ted-13	ted-14	Average
Arabic→English					
Phrase-based	30.5	34.2	35.0	30.5	32.6
Neural	32.5	37.0	37.2	31.5	34.6
SysComb	32.8	36.5	37.4	31.7	34.6
English→Arabic					
Phrase-based	16.7	17.9	20.2	17.7	18.1
Neural	17.1	18.9	20.1	17.7	18.5
SysComb	16.8	19.1	20.7	17.6	18.6

System	ted-15	ted-16	qed-16
Arabic→English			
Primary	34.1	31.8	28.1
Contrast	33.7	31.5	28.1
English→Arabic			
Primary	19.5	18.4	23.1
Contrast	19.5	18.1	22.9

*All scores shown are cased BLEU scores. English→Arabic system outputs were detokenized using MADA detokenizer, and normalized using QCRI normalizer.

ACKNOWLEDGEMENTS

We would like to thank:

- Texas A&M Qatar for providing computational support
- University of Edinburgh for their support during this competition