From statistical machine translation to translation with neural networks for the Slovene-English language pair

Gregor Donaj, Mirjam Sepesy Maučec

University of Maribor
Faculty of Electrical Engineering and Computer Science
Digital Signal Processing Laboratory

September 2018

Machine translation

- Automatic translation from one natural language to another.
- Focus on approaches that learn translation principles from bilingual (and monolingual) text corpora.
- For a long time SMT brought the best results.
- Recently NMT emerged with the promising results that outperformed SMT.

Statistical MT

- A machine translation paradigm based on noisy channel model.
- Independently learned components: translation model (TM), language model (LM), reordering model (RM).
- Phrase-based SMT: nuclear unit is a data-driven phrase.
- Component weights are determined using minimum error rate training.

Neural MT

- One large model with no explicit division into TM, LM and RM.
- Recurrent neural network (RNN) works on a variable-length sequence.
- Attentional encoder-decoder architecture.
- Attentional mechanism implies a set of source hidden states which are consulted during translation process.

Our goals

- First steps in neural machine translation with Slovene in the language pair.
- Comparison of results between NMT and SMT.
- Assessment of tools and results.
- Establish a baseline system and model for further research.

Training and test corpos

- Europarl (version 7).
- Transcription from the European parliament and translations.
- Training, development and evaluation sets.
- cca. 600.000 lines of text and 12/15 mio. words.
- 2000 lines for development and evaluation.

NMT Tools

- Marian (former AmuNMT) for training and translation.
- Models in Nematus format.
- Moses for (de)tokenization and (de)normaliaztion.
- Multeval for evaluation (BLEU, TER and METEOR).

SMT Tools

- Moses 3.
- SRILM for language models.
- MERT for optimization.
- For (de)tokenization, (de)normaliaztion, and evaluation: same tools and models as in NMT.

Training speed

- Speed depends on model and training hyperparameters. Following numbers are for default parameters of Marian.
- Training speed: 368 (sl-en) and 390 (en-sl) sentences per second.
- For the Europarl corpus, this is 2.5 or 2.3 epochs per hour.
- Total training in our experiments was 22,5 epochs in each direction, requiring a total amount of 18 hours of training (on a Nvidia 1080 Ti).

Hyperparameter optimization

- Optimal training duration: 110.000 iterations (mini-batches of 64 sentences) = cca. 12 epochs.
- Mini-batch size: 64. Results with mini batches between 16 and 256 are lower by up to 2 point BLEU.
- Optimal embedded vector size: 512.
- Optimal RNN size: 1024.
- Results for non-optimal values drop quicker in direction sl-en.

System comparison

Table : Translation evaluation results for the NMT system and the comparison to the SMT system (Δ).

	En→Sl	Sl→En
BLEU	40,8	46,4
METEOR	33,2	41,7
TER	43,0	37,1
Δ BLEU	5,3	2,3
Δ METEOR	2,1	-0,1
Δ TER	-3,0	-1,8

Out of domain results

Table : Out of domain translation results: Orwell, 1984.

	En→Sl	Sl→En
SMT	9,0	10,5
NMT	8,5	10,0
Δ	0,5	0,5

Conclusions

- We got better results with NMT.
- Better performance improvement in the direction from English to Slovene.
- Optimal training duration with a similar number of epochs when comparing to other research.
- Further work: inclusion of morpho-syntactical information, including a mono-lingual language model for hypotheses re-scoring, new domain adaptation.

Thank you for your attention!

Any questions?