QuidditchChampions

Markus Kängsepp, Ian Mackerracher, Andre Tättar
University of Tartu, Institute of Computer Science

Study IT in .ee
sponsored by Skype™

Introduction

Within the scope of the course we used OpenNMT-py to build a machine translation model that would translate text from Estonian to English. We started off with a baseline model, which got a BLEU score of 21.95. Our goal was to make modifications to the baseline model and improve the translations and ultimately the BLEU score. With our final model we achieved a score improvement of 22.89 BLEU points on the Dev Accurate set.

We explored a number of approaches in order to improve our score. Ranging from replacing dots within sentences to counter unwanted splitting of sentences, employing coverage and context gates, varying beam sizes, hyper-parameter tuning and varying the size of the BPE vocabulary.

For our final model we used Amazon's Sockeye library to train a system using context gates and instead of attention we used coverage, our BPE vocabulary size was 30k. For translation we used a beam size of 10.

Things We Tried

- Dot-model - replacing dots that appear in the middle of sentence with some other token.
- Beam sizes - bigger beam sizes slightly improved performance, since it saw more translation data.
- Vocabulary size - hoped that bigger vocabulary makes translation better.
- Hyper-parameter tuning - took parameters from Google Brain's paper, which resulted in good converging times and good results.
- Context gates - dynamically control the ratios at which source and target contexts contribute to the generation of target words.
- Coverage - indicates whether the source word has been translated or not. Replaces attention.

Examples

Example 1

Reference: The government is formulating an efficient healthcare system based on a long-term perspective.
Baseline: The government will develop an effective health system based on long-term prospects.
Final Model: The government is developing an effective health care system based on a long-term perspective.

Example 2

Reference: 6. publishing our own information.
Baseline: 6.
Final Model: 6. publication of your information.

Example 3

Reference: The possible changes of the department organisation are just now under consideration.
Baseline: Possible changes in the department's organisation are already at stake.
Final Model: Possible changes in the organisation of the department are already under consideration.

Example 4

Reference: Suddenly you could listen to another person's voice hundreds of miles away.
Baseline: Suddenly, the voice of another person has been heard from hundreds of miles.
Final Model: Suddenly, you could listen to another person's voice from hundreds of miles away.

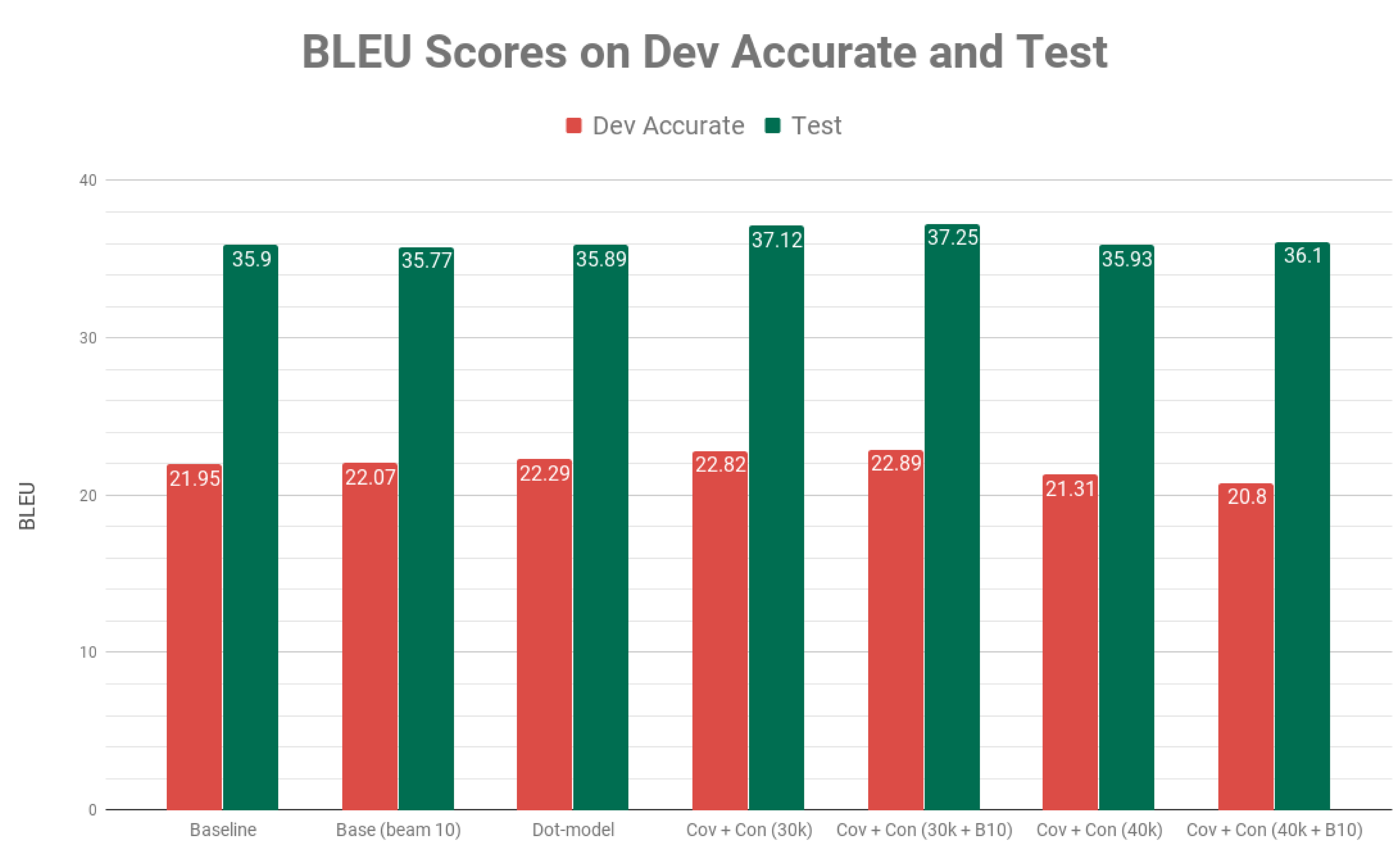

Figure 1: BLEU scores of Dev Accurate set and Test set with different models and beam sizes

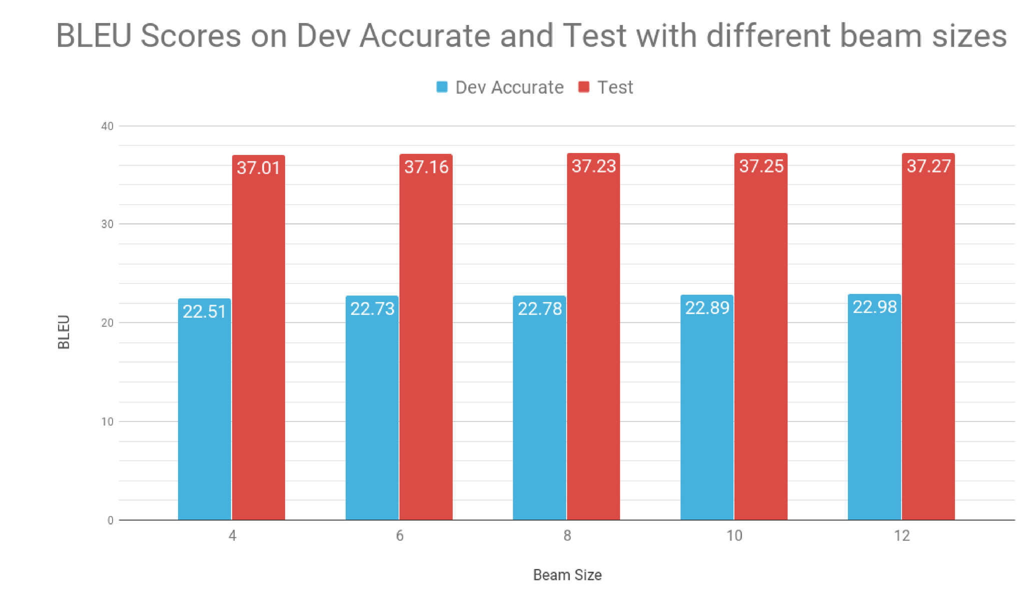

Figure 2: BLEU scores of Dev Accurate set and Test set with different beam sizes

Conclusion

We explored a number of options to improve the translation quality of the system. In the end we found that using Amazon's Sockeye library to train the system yielded the best results. For this we used context gates and instead of attention we used coverage. Furthermore we used a BPE vocabulary size of 30k and a beam size of 10. We found that this produced the best results, both in terms of manual inspection and BLEU score.

Discussion

- We learned that there is no guarantee with neural networks. Sometimes it does not work and you have no idea what is going on. We thought we picked the best parameters from Google Brain's paper, but it performed horribly.
- Sockeye has some significant advantages over OpenNMT. With OpenNMT, on the same hardware it trained 7 days, but Sockeye converged and finished training within less than 30 hours.