Lab 4: Improving NMT

by Maksym Del and Mark Fishel

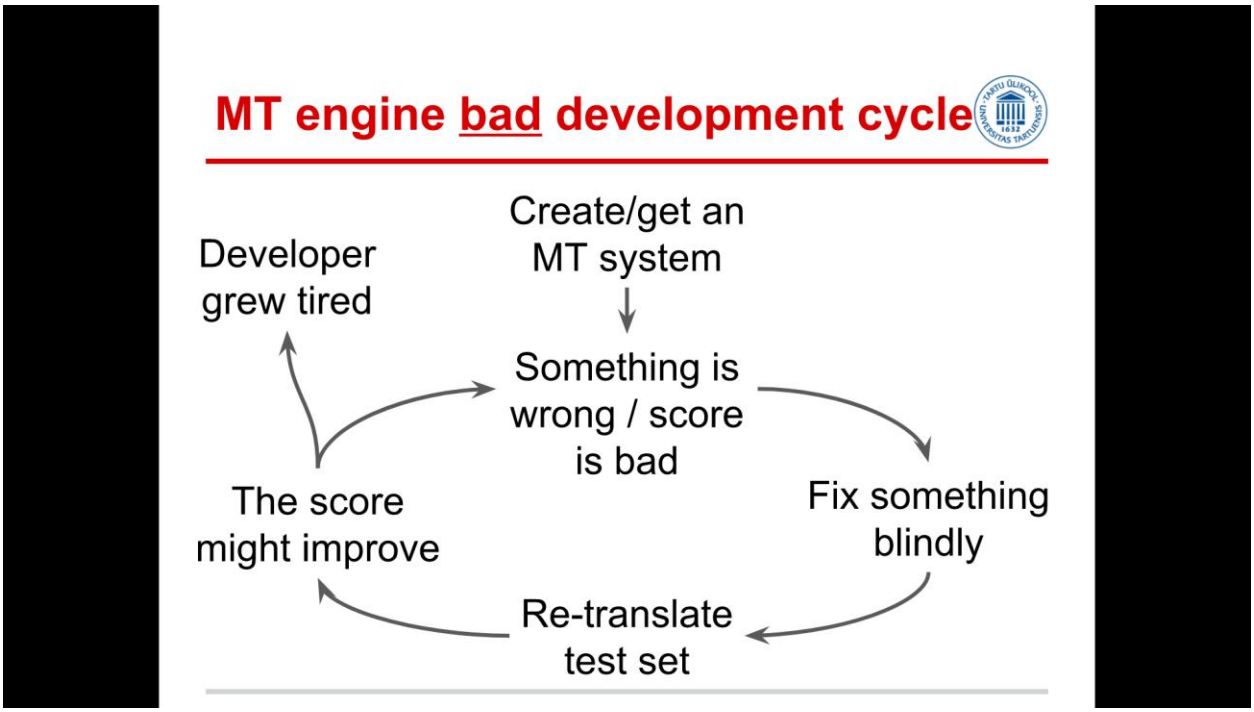
Introduction

By this point, you got necessary skills needed in order to build and analyze modern NMT systems. You were guided through the MT systems developing workflow, and now it is time for you to continue almost by yourself. You have your baseline trained and saw what kind of mistakes the baseline system does. During the rest of the course you will try to improve on initial system in terms manual evaluation of BLEU.

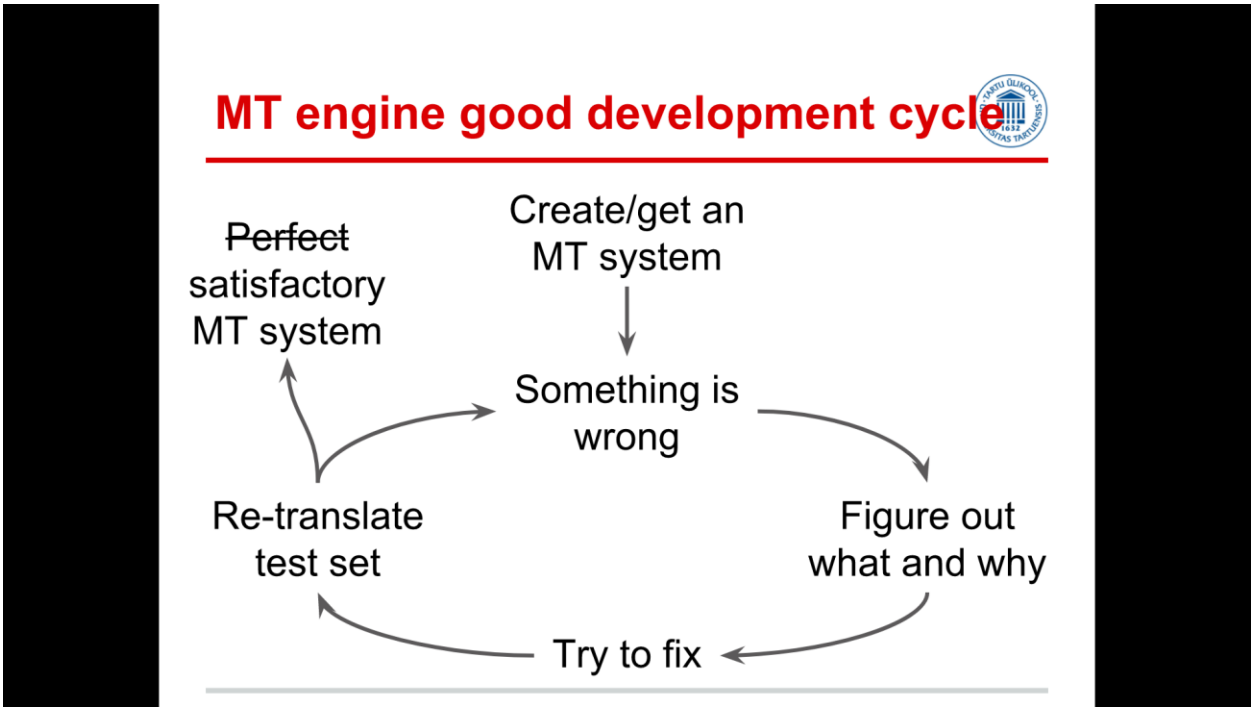
At this lab we look at practical aspects of some initial on MT improving you might employ.

Good and bad MT engine development cycle

First, let me remind you how good and bad development cycles look like (slides are from the course lecture on error analysis):



MT engine **bad** development cycle



Mt engine **good** development cycle

The key message to be taken is that your improvements should be guided by results of the system analysis. You had a chance to try it out during previous lab and milestone. So the next step is to incorporate some fix to the MT system based on the intuition you got from the error analysis of your baseline.

Report on system modification

At home, you have to turn your manual analysis work results into some concrete problem your baseline system has, and then link this problem with one of the modification from the list below.

In the report, you should describe the problem you defined (with concrete sentence examples from your manual evaluation milestone), the modification you chose (with some theoretical foundations; see “example NMT improvements” chapter of this document), and give concrete explanation of how you think the modification is going to help with the problem.

Moreover, you have to fill your project board (on GitHub) with all the steps (tasks) you are going to do in order to actually incorporate the modification in your system. Include also approximate dates for the steps in your report.

You can try on or more modifications; however, it also depends on the difficulty of the fix, and time needed to incorporate it into the system. So be careful with your time, since you have 3 weeks for the first iteration. Next, you will have had one more iteration before the test set input-side is released.

Example NMT improvements

One main resource you can (and should) look first is the NMT chapter for the textbook Statistical Machine Translation written by Philipp Koehn (https://arxiv.org/abs/1709.07809). However, we also provide you some papers to dive deeper. You need the theoretical info on modification planned because you should describe it in your report.

List of possible modifications:

- Try fully convolutional sequence-to-sequence models
  - Soft: 1, 2
  - Papers: 1
- Try Transformer Models with self-attention
  - Soft: 1, 2
  - Papers: 2
- Translate long sentences by chunks
  - Soft: 2
- Try model ensembling or/and checkpoint averaging:
  - Soft: 2
  - Papers: 3
- Hyperparameters tuning (incl. deep architectures, and/or weight/layer normalization, and/or dropout, and/or optimizer type, …):
  - Soft: 1 / 2
  - Papers: 4, ...
- Context gate + Coverage:
  - Soft: 1 / 2

- Papers: 5, 6

- Incorporating linguistic info (e.g. apply POS tag the input): direct pre-processing / via word features:
  - Soft: 1 / 2
  - Papers: 7, ...
- Post-process data:
  - Soft: 1 / 2
  - Papers: ...
- Use more data via back translation and/or paraphrasing:
  - Soft: 1 / 2
  - Papers: 9, ...
- Try pretrained embeddings (tune them or fix):
  - Soft: 1 / 2
  - Papers: 9
- Go fully/partially character level without/with explicit segmentation (char2char / bpe2char / char2bpe):
  - Soft: 1 / 2
  - Papers: 10, ...
- Replace BPE with “wordpieces”, or try other fixes to subword unit generation -- if subword units are often bad:
  - Soft: 3 + 1 / 2
  - Papers: see soft
- Implement recent “Swish” activation function:
  - Soft: -
  - Papers: 11
- Your idea

Soft:

1 - OpenNMT-py: https://github.com/OpenNMT/OpenNMT-py

2 - Sockeye: https://github.com/awslabs/sockeye

3 - Google's sentencepiece: https://github.com/google/sentencepiece

Papers:

1 - https://arxiv.org/abs/1705.03122

2 - https://arxiv.org/abs/1706.03762

3 - https://arxiv.org/abs/1708.00726

4 - https://arxiv.org/abs/1703.03906

5 - https://arxiv.org/abs/1601.04811

6 - https://arxiv.org/abs/1608.06043

7 - http://www.statmt.org/wmt16/pdf/W16-2209.pdf

8 - https://arxiv.org/pdf/1511.06709.pdf

9 - https://arxiv.org/abs/1301.3781

10 - https://arxiv.org/abs/1610.03017

11 - https://arxiv.org/pdf/1710.05941.pdf