In-class Machine Translation Shared Task 2017

Part of the MTAT.06.055: Machine Translation course

Organized by Maksym Del & Mark Fishel

Language pair:
Estonian -> English

1 Stage I: Neural Baseline

Training data is released: 19,000,000 sentence pairs

Teams split and preprocess the raw data

Teams train baseline system: neural seq2seq model

Teams use MT best practices and cluster computing

Teams use GitHub for project management and VC

2 Stage II: Manual Error Analysis

In-domain dev set is released

Teams translate dev set with their baselines

Teams manually analyze 40-60 sentences and define main error types

Teams inspect attention weights (internal features) of the NMT models

6 teams
6 NMT systems
5 milestones
18 weeks
1 test set
1 winner

3 Stage III: Improved Baseline System

Teams independently work on modifications to the usual NMT baseline to address issues based on manual error analysis

Teams translate in-domain dev set with their modified systems

Teams manually analyze the effect of their changes

4 Stage VI: Final systems and submissions

Teams continue working on ways to improve their systems

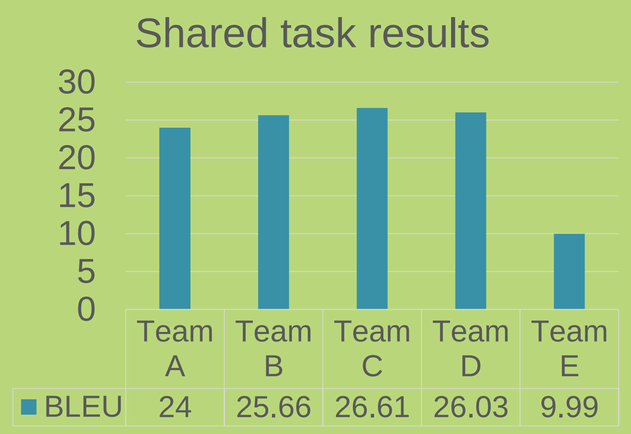
Teams train final systems to make submission

Source side of the test set is released

Teams submit final translations of the test set with their final systems

Machine translation!
Make systems
that speak better
then Yoda

5 Final results



Each team prepared its own poster!

Come to learn their stories!

ATTENTION!!! ATTENTION!!! ATTENTION!!!

What is next? WMT 2018!

We all become a single team to take part in the international machine translation shared task.

Register to participate:
Seminar on Language Technology (MTAT.06.046)

* Links

1. Contact: maksym.del@gmail.com | fishel@ut.ee
2. Course page: https://goo.gl/v7ij1a
3. Shared task page: https://goo.gl/3vY1x7
4. Picture source: https://goo.gl/mi9rAB

Acknowledgement:
UT HPC center for GPUs and support