Solving Three Czech NLP Tasks End to End... ... with Neural Models

Jindřich Libovický, Rudolf Rosa, Jindřich Helcl, and Martin Popel

Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics

Image Captioning

- Architecture "Show, attend and tell"
- Czech part of the Multi30k dataset for multimodal translation
- First image captioning model for the Czech language (although the scores are rather poor)

Machine Translation

- Cutting-edge transformer architecture
- Trained on CzEng 1.7, a parallel corpus with 57M sentence pairs
- State-of-the-art results on English-to-Czech WMT news translation shared task

Sentiment Analysis

- Bidirectional LSTM with self-attentive layers
- CSFD CZ dataset, 91k movie reviews from the Czechoslovak film database
- Competitive results, though not outperforming the current state of the art

http://github.com/ufal/neuralmonkey

http://quest.ms.mff.cuni.cz/neuralmonkey-czm

This poster was presented on September 22, 2018 at SloNLP in Krompachy, Slovakia. This work has been supported by the grant No. 18-02196S of the Grant Agency of the Czech Republic, the grant No. CZ.02.1.01/0.0/0.0/16_013/0001781 of the Ministry of Education, Youth and Sports of the Czech Republic, the grant No. 976518 of the Grant Agency of the Charles University, and the SVV 260453 grant. This work has been using language resources and tools developed, stored and distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LIIAT753).