

Uladzislau Yorsh

Teplická 604/15
190 00 Praha 9, Praha
☎ +420 777 162 444
✉ vladyorsh@gmail.com

Education

- 2026 (exp.) **Charles University in Prague**, PhD.
2022 **Czech technical university in Prague**, Ing, with honors.
2020 **Czech technical university in Prague**, Bc.

Experience

- Apr 2021–Oct 2022 **Research Assistant**, THE BIGCODE PROJECT.
2022
 - Proposed, implemented and evaluated two architectures for processing sequential inputs with $\mathcal{O}(n)$ complexity w.r.t. a sequence length.
- May 2021–Oct 2021 **Data Science Intern**, RECOMBEE S.R.O.
2021
 - Proposed and developed a new model for the next basket prediction task, which improved the IoU score by 5%.
 - Used a LSTM which incorporated a suitable "last matters more" inductive bias.
 - 8M interactions dataset; Tensorflow implementation.
- Mar 2021–May 2021 **Research Assistant**, INFERENCE TECHNOLOGIES.
2021
 - Proposed and implemented an unsupervised classification algorithm of wafer bin map defects
 - Improved the existing system performance from 0.76 to 0.81 kappa score.
 - Used an autoencoder to embed WBM into a latent space, a denoising variant to make a more robust embedding for classes with defects looking similar to noise.
 - Additionally experimented with several variational and adversarial autoencoder variants.

Publications

- ICANN 2022 **Linear Self-Attention Approximation via Trainable Feedforward Kernel.**
 - Proposed and implemented a new attention mechanism with a linear complexity w.r.t. an input sequence length.
 - Evaluated the model on the LRA benchmark and beaten most of the baseline models.
- ITAT 2022 **Text-to-Ontology Mapping via Natural Language Processing Models.**
 - Explored the possibilities of an automatic assignment of an ontology to a text document.

Other Projects

- Aug 2021–Feb 2022 **SimpleTRON: Simple Transformer with $\mathcal{O}(N)$ Complexity.**
 - Proposed and implemented a new attention mechanism with a linear complexity w.r.t. an input sequence length.
 - Evaluated the model on the LRA subset and outperformed all other models on the considered tasks at the moment of publication.
- Feb 2023–present **Shared Task on Automatic Minuting.**
 - Working on adopting one of the sub-quadratic Transformer variants for summarization of long texts.

Skills

- Languages Python, C, C++, Scala, Java, JavaScript, R
- Domains Machine Learning, Computer Vision, Signal Processing, Natural Language Processing, Data Preprocessing, Theoretical Informatics, Statistics
- Technologies PyTorch, Tensorflow, Keras, JAX, MATLAB, SQL, Docker, Apache Cassandra, Elasticsearch, MongoDB, Hadoop, SPARK, git, REST
- Communication English (B2), Czech (B2), Russian (native)

Research Interests

Handling long sequences, which contain tens and potentially hundreds of thousands of tokens, is a challenging machine learning task I tackle. The goal is to make the models to be able to work with text chapters or even whole articles, in contrary to the current state-of-the-art limitation of few paragraphs. As this task is highly hardware demanding, I am also keen on developing parameter-efficient models that are able to reach state-of-the-art performance using less amount of compute, and put an additional effort into backing my findings up with theoretical explanations.