mihlushi@yandex.ru
https://github.com/MichaelSolotky
https://www.linkedin.com/in/michael-solotky/

Michael Solotky

Moscow, Russia
+7 967 291-08-66

Education

- MSc student in Applied Mathematics and Computer Science
  Higher School of Economics : Faculty of Computer Science September 2019 – June 2021
  Joint programme with Yandex School of Data Science
- BSc in Applied Mathematics and Computer Science, GPA 4.89 / 5.0
  Lomonosov Moscow State University September 2015 – June 2019
  Faculty of Computational Mathematics and Cybernetics

Continuing with a PhD degree afterwards

Experience

- **Machine Translation department of Yandex** June 2019 – September 2019
  Software Engineering Intern (Machine learning engineer)
  - Conducted experiments to improve quality and diversity of translations
  - Analyzed and found some basic mistakes that baseline approaches do
  - Implemented several successful models and inference techniques in the Yandex's machine learning library with an ability to control diversity level
  - **Achieved statistically significant improvement in quality and diversity simultaneously** on Yandex's metrics, commonly used in scientific field metrics (such as BLEU and self-BLEU) and human evaluation compared to the baselines
- **Voice Technology department of Yandex** June 2018 – October 2018
  Software Engineering Intern (Back-end developer)
  - Implemented several methods of probability smoothing and their modification in language models for Automatic Speech Recognition
  - Conducted experiments on quality measurement to find the best model among all
  - Implemented an optimal algorithm for constructing n-gram language models in C++ using MapReduce, **which decreased wall time by at least 3 times and slightly increased quality measure** compared to baseline
  - Wrote a complete framework with a set of operations available from CLI

Projects

- **BigARTM (C++ Boost/STL, Protobuf, Travis, AppVeyor)** January 2017 – May 2019
  Open Source library for topic modelling with support of multiple regularization
  github.com/bigartm/bigartm
  - Developed and supported a tool for parallel calculation of pairwise word statistics such as frequency of mutual occurrence, PMI in large text corpora in conditions of low RAM
    Wikipedia full-text processing takes 6 hours on octa-core intel core i5 8th gen, taking less then 8 Gb of RAM

Other experience

- **ML (NumPy, Scipy)**
  Implementation of various ML algorithms from scratch
  github.com/MichaelSolotky/sandbox/tree/master/Machine_Learning

Technical Skills

- **Languages used at work**: C++, Python, C, Bash
- **Basic knowledge**: SQL, Assembly language
- **Technologies**: MapReduce, Protobuf, C++ Boost, CMake, Make, SciPy, Scikit-learn, NumPy, Pandas
- **Deep Learning frameworks used at work**: PyTorch, TensorFlow
- **Tools**: Git, Subversion, UNIX/Linux, Travis, AppVeyor, LaTeX