 GitHub: MichaelSolotky

 LinkedIn: michael-solotky

Michael Solotky

Moscow, Russia

✉ mihlushimihlushi@gmail.com

EDUCATION

- MSc student in Applied Mathematics and Informatics, GPA 9.23 / 10.0
  Higher School of Economics : Faculty of Computer Science   Sep 2019 – Jun 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   Sep 2015 – Jun 2019
  Faculty of Computational Mathematics and Cybernetics

EXPERIENCE

- **Research Intern at Yandex (PyTorch)**   Sep 2020 – now
  - Finding **theoretical foundation** for different methods of **uncertainty estimation** in **Deep Learning**
  - **Performance validation** of existing methods for **uncertainty estimation** in different environments
- **Software Engineering Intern at Yandex**
  Machine Translation department (TensorFlow, MapReduce, NumPy, SciPy)   Jun 2019 – Sep 2019
  - Conducted experiments to improve quality and diversity of translations
  - Analyzed and found some basic mistakes that baseline approaches make
  - 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, human evaluation compared to the baselines and commonly used in scientific field metrics: **about 1.1 growth of max-BLEU** (maximum BLEU of generated translations) and **about 1.6 times n-gram diversity** (1 - self-BLEU) **growth**
- **Software Engineering Intern at Yandex**
  Voice Technology department (C++ STL, MapReduce, Protobuf)   Jun 2018 – Oct 2018
  - 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

- **Actual AI Designer (PyTorch)** – Mentorship in a school project   Aug 2020 – now
  A tool for automatic generating logos based on the description (in the very beginning)
  - **Teaching** students the necessary topics in **machine learning, computer vision** and other technical skills sufficient to complete the project
- **BigARTM (C++ Boost/STL, Protobuf, Travis CI, AppVeyor)**   Jan 2017 – May 2019
  Open Source library for topic modelling with support of multiple regularization
  - 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 than 8 Gb of RAM compared to at least 20 Gb needed before (code sample)
- **Various Contributions:**
  - Support TensorFlow 2.x versions of home tasks of the Yandex's **Reinforcement Learning course**
  - Answering questions on **stack exchange** (sometimes)

TECHNICAL SKILLS

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