Shamil Iakubov : LinkedIn Email : yakubov.sha@gmail.com

Machine Learning projects

- **Top 2% solution in the Featured Simulation Kaggle Competition "Lux AI"**: Ranked 19th out of 1186 teams as a solo participant (team shmyak). The solution is based on imitation and distributed reinforcement learning. It consists of:

  1. Environment. An OpenAI gym wrapper for a Kaggle environment, which does preprocessing of raw data to provide ready observations to a trainer agent and contains several rule based agents. Several workers use the environment to collect experience to a data buffer.

  2. Trainer. It performs training of a function approximator, it includes different implementations of actor-critic and policy gradient algorithms, a custom training loop for imitation learning, and a custom buffer to store game experience. The buffer uses tfrecords files to prevent storing all experience trajectories in memory but consuming them efficiently from a storage device. It uses EfficientNetV2 squeeze-and-excitation layers as a function approximator.

- **Developed a policy gradient based reinforcement learning algorithm for a custom environment**: The implementation is close to IMPALA. It includes a custom training loop of an off-policy actor-critic algorithm with n-step update, policy gradients correction, entropy, and other improvements. It uses different convolutional neural nets as a function approximator and applies multi-attention for data preprocessing.

- **Developed a DQN based reinforcement learning algorithms**: The project inludes custom training loops implementation of several reinforcement algorithms (TensorFlow): DQN, FixedDQN, DoubleDQN, DoubleDuelingDQN, categorical DQN. It uses RAY to distribute calculations and DM Reverb as a data buffer to perform Prioritized Experience Replay. It uses a sparse MLP as a function approximator.

- **Conducted an optimization study**: The research involved developing a model that predicts the ocean's ability to absorb carbon dioxide. It uses a custom function approximator and a non-linear least-squares minimization.

- **Conducted a signal processing study**: It proposes the method to predict 'Freak waves' based on waves parameters. The study uses cluster analysis to categorize waves to groups and then uses Fourier and Wavelet analysis to study properties and features of these groups.

Related experience

- **Helmholtz-Zentrum in Geesthacht** Geesthacht, Germany
  PhD student *2017 - 2020*
  - **Research**: Study of biogeochemical interactions between the ocean and the atmosphere in the North Sea region. Writing and publishing scientific papers. Processing and analyzing geospatial data using Pandas, Numpy, Matplotlib, etc.
  - **Development**: Building and optimization of ocean ecosystem and biogeochemical models.

- **Institute of Oceanology** Moscow, Russia
  Research engineer *2014 - 2017*
  - **Research**: Processing and analysis of ocean waves. Studying Arctic ecosystems and biogeochemical processes.
  - **Development**: Introducing the CMake build system to multiple Fortran projects. Migration from Fortran90 to modern object-oriented Fortran. Development of a diffusion-advection model of particle transport in the ocean.

Programming Skills

- **Languages**: Python (Tensorflow, Keras, Numpy, Pandas), FORTRAN, LaTeX

Recent publications

- Yakubov, S.; Protsenko, E. Alkalinity Generation in the Coastal Area, the Case of the Wadden Sea. Preprints 2021, 2021020036 (doi:10.20944/preprints202102.0036.v1)

Education

- **Moscow State University** Moscow, Russia
  Specialist, Oceanography *2003 – 2008*

Online courses: Deep Learning specialization, Machine Learning, Bayesian Statistics.