Ruize (Richard) Luo

1609 - 15 Greenview Ave, North York, Ontario, Canada, M2M4M7

+1 (647) 339-5825 | ruize.luo@outlook.com | r-luo | luoruize

Skills

Languages: Python (PyTorch, numpy/pandas, scikit, dask, SimPy), R (data.table, ggplot2), SQL, Bash, SAS, VBA
Platforms: Tableau, AWS (EC2, S3, RedShift), Relational Databases (Teradata, Snowflake), *nix
Machine Learning: Gradient Boosting Machines (GBM), Large-scale Feature Engineering, Bayesian Model Selection, Feature Selection, Model Pipeline, Model Interpretation, Reinforcement Learning, Neural Networks (Autoencoder, CNN, GAN), Kalman and Particle Filters, AI Search and Planning, Optimization, Semi-supervised Learning

Work Experiences

Amazon
Toronto, Ontario, Canada

Data Scientist II | Sponsored Product Advertising
Jul. 2020 - Present

- Immediately delivered value by building a pipeline that streamlines data processing, model build and validation to accelerate the worldwide expansion of my project from the planned 3 months to 3 weeks

Capital One Bank
Toronto, Ontario, Canada

Principal Data Scientist | Fraud Team
Jun. 2018 - Jun 2020

- Leading data science projects for application fraud.
  - Migrated current model from legacy platform to Python on AWS as part of our enterprise-wide cloud migration.
  - Refitted the model with LightGBM. Independently finished model build from data collection to deployment, designed the downstream business policy and provided valuation for model impact on business budgeting. The refit reduced expected loss by 40% and false positive rate by 20%.
  - Utilized hyperparameter tuning and feature selection techniques to improve model performance. Developed a Markov Chain Monte Carlo based hyperparameter search algorithm which avoids the overhead of popular Bayesian based hyperparameter optimization techniques for smaller sized problems.
  - Explored semi-supervised learning methods to utilize unlabeled data. As a result, our models adapt better to recent trends in the population.
  - Supervised a co-op student on the deployment of a real time model on our streaming platform.
  - Supervised a second co-op student to modernize our fraud defence valuation pipeline. It reduced the time needed for developing new fraud rules from weeks to hours. Received great feedbacks from both coops.
  - Prototyped a data pipeline as a python module to bring model development and deployment into one place.
- Started a journal club within the data scientist community, facilitated bi-weekly paper discussions to bring the team up-to-date with the academia, and to drive team engagement and innovation.
- Promoted to Principal Data Scientist in July 2019.

Senior Data Scientist | Customer Management Team *Sep. 2015 - May. 2018*

- Built a core business model and its end-to-end scoring pipeline for predicting customer default.
  - The whole process involved sample selection, data pull, data cleaning and validation, feature engineering and selection, model build and validation, model deployment, documentation, and ongoing model monitoring.
  - Worked with datasets on the magnitude of 100GB.
  - Used SQL, SAS and Tableau for data cleaning and validation; used GBM and Bayesian model selection techniques in R for feature selection; explored genetic algorithm along the way.
  - With careful data engineering (various variance stabilizing transformations, and splines) and model selection, our logistic regression model was able to achieve similar performance as tree-based GBMs while being much more interpretable and stable over distributional shifts
  - Deployed the model on our internal scoring platform as a Python package.
  - Guided by Agile principles: fast iterations of minimum viable product, quick adaptation to changes, and integration with business team for smooth communication.
- Independently developed a Python package to perform large-scale feature engineering in parallel with Dask. The package employs a graphical representation of feature lineages to allow easy parallelization and extraction of feature lineages. The package is now used across multiple teams to standardize feature generation and documentation.
- Proactively helped operations team to build a Monte Carlo simulation tool in Python to simulate customer call queue. It significantly improved call centre staffing to shorten customer wait time and reduce operations cost.
- Promoted to Senior Data Scientist in July 2017.

Projects

CS7642: Reinforcement Learning, Georga Institute of Technology *Toronto, Ontario, Canada*

Course Project *Mar 2019*

- Built a deep reinforcement agent to solve the OpenAI Gym Lunar Lander environment and successfully solved the environment by achieving a score greater than 200 over 100 trials.
- Utilized double deep Q network with experience replay and explored the effect of different hyperparameters on training.
- Source code and project report available upon request.

Bone X-Ray Deep Learning Competition from Stanford ML Group *Toronto, Ontario, Canada*

AI Squared Forum *Jun 2018 - Oct 2018*

- The challenge is to determine whether the bones presented in the x-ray pictures are normal. There are ∼38K images from ∼15K different studies.
- Led a team to learn basic computer vision techniques by participating in the challenge.
- Set up data pre-processing pipelines, training pipelines and designed evaluation metrics with Keras. Achieved result on par with the benchmark model with an ensemble of DenseNet trained with Adam using cyclical learning rate.
- **Github**: DeepMachineLearning/mura-team2

Education

Georgia Institute of Technology *Online*

Masters of Science, Computer Science *Jan. 2019 - Present*

GPA: 4.0/4.0

- CS7642: Reinforcement Learning (A)
- CS7638: Artificial Intelligence for Robotics (A)
- CS7646: Machine Learning for Trading (A)

University of Toronto *Toronto, Ontario, Canada*

Non-Degree Graduate Program, Computer Science *Sep. 2017 - Dec 2017*

GPA: 4.0/4.0

- CSC2420: Algorithm Design, Analysis and Theory (A+)
- CSC2221: Introduction to the Theory of Distributed Computing (A+)

University of Toronto *Toronto, Ontario, Canada*

Honours Bachelor of Science, Statistical Sciences Specialist and Actuarial Science Major *Sep. 2011 - Jun. 2015*

CGPA: 3.94/4.00 (Major Course Average: 95.4/100.0)

- Courses: Data Mining, Statistcal Computation, Time Series, Stochastc Processes, Loss Models, Real Analysis, Financial Mathematics

Hobbies

Performance driving, reading, photography, badminton, foosball, gaming, hiking, biking