Shuto Araki

GitHub : https://github.com/ShutoAraki

Email : shutoaraki_2020@depauw.edu

Mobile : +1-765-301-1185

Education

- **DePauw University** — Greencastle IN, USA
  Bachelor of Arts, double major in Computer Science and Mathematics; minor in Economics — *Aug. 2016 – May. 2020*
  - **GPA**: 3.96/4.00 cumulative, 4.00/4.00 in both majors
  - **Awards**: Rector Scholarship (4 year full-tuition), Wylie Condit Computer Science award, Phi Beta Kappa
  - **Activities**: First-Year Resident Assistant, Information Technology Associates Program, Science Research Fellows, Co-president of DePauw Data Science Group, Delta Upsilon Fraternity, Analytics editor at The DePauw
  - **Work Authorization**: F-1 visa status - I am authorized to work in the U.S. without sponsorship until July 2023.
  - **Relevant Coursework**: Data Mining, Artificial Intelligence, Compilers, Databases & File Systems, Graphics, Linear Algebra, Graph Theory, Operations Research, Probability, Financial Engineering

Experience

- **Oak Ridge National Laboratory** — Oak Ridge TN, USA
  Advanced Data Workflow and Analytics Intern — *June 2019 - August 2019*
  - **Distributed Deep Learning**: Trained the ResNet-50 on the whole ImageNet dataset in under 2 minutes from scratch (state-of-the-art result) using the world's fastest supercomputer, Summit. Challenge was to fully utilize the 27,600 V100 GPUs. Achieved faster convergence of the ResNet-50 at scale from two approaches: efficient data I/O and communication among 4,600 nodes through TFRecord format and memory optimization such as mixed precision training, deep gradient compression, and Layer-wise Adaptive Rate Scaling (LARS).
  - **Technologies and Skills**: Linux, Bash, JupyterLab, TensorFlow, Mesh-TensorFlow, PyTorch, Keras, Horovod
- **DePauw University** — Greencastle IN, USA
  Summer Researcher — *May 2018 - August 2018*
  - **Research**: Created highly optimized algorithms to determine solvability of a card game invented for a research competition using Machine Learning regressors and graph theoretic tree pruning techniques. Worked with four other students and professors to explore and develop novel algorithms without any previous papers on the card game.
  - **Technologies and Skills**: LightGBM, Graph theoretic feature engineering, Scikit-learn, Pandas, cache optimization on Python, object oriented programming in Python and Java, Git
  - **Publication**: Published and presented the paper at the AAAI 2019 international conference in Honolulu. Won the second best paper award.
- **Coca-Cola (Japan) Company, Ltd.** — Tokyo, Japan
  Summer IT intern — *May 2017 - August 2017*
  - **Data Pipeline Automation**: Wrote a script that automates daily data transfer process from Amazon AWS S3 to Google BigQuery using Google Cloud Functions. Challenge was to learn all the technologies in three weeks and automate the pipeline. This system shortened the data processing time from one human hour to two minutes and eliminated human errors.
  - **iPhone App Feature Development**: Updated an iPhone app with a vending machine analytics team. Worked on developing and testing new features for their campaigns. The app had over 7 million downloads across Japan.
  - **Technologies and Skills**: Google BigQuery, SQL, Excel, VBA, application of technology in business context, execution of plans by communicating with other Coca-Cola employees across the world in Japanese and English.

Projects

- **Universal Crystallography Classifier**: Team project. Designed a data pipeline that converts 580GB of HDF-5 format data into TFRecord (serealization) format and PyTorch DataLoader. This enabled us to feed all the data into a model on TensorFlow/PyTorch efficiently without worrying about memory overflow.
- **Foreign Exchange Rate Forecast Model**: Implemented a hybrid model of ARIMA (paprametric) and modified k Nearest Neighbors in time series (non-parametric) that can capture non-linearities in foreign exchange rate data. Explored error surface of the data in addition to hyperparameter tuning.
- **Rare Genetic Disease Mutation Highlighter**: Enhanced research productivity in the biochemistry department by creating a custom extension on the UCSF Chimera visualization software. Learned to navigate through poorly documented code base (developed by biochemists) with Object-Oriented Principles in Python.

Additional Technical Skills

- **Languages**: Python, Java, C++ **Technologies**: Bash, Git, Docker, CircleCI, Angular, Django2.x