Yu Xiang

288 106th Ave NE, Bellevue, WA 98004

📞 518-833-4710 ✉ yux1991@gmail.com 🔗 linkedin.com/in/yu-xiang-physics ⦿ github.com/yux1991

EDUCATION

Doctor of Philosophy in Physics, **Rensselaer Polytechnic Institute**　　　　　　**Aug. 2014 – Dec. 2019**
GPA: *3.88* of *4.00* | Relevant Courses: *Quantum Mechanics, Statistical Mechanics*　　　　　Troy, NY

Bachelor of Science in Physics, **Wuhan University**　　　　　　　　　　　　**Sep. 2010 – Jun. 2014**
GPA: *3.43* of *4.00* | Relevant Courses: *Advanced Mathematics, Linear Algebra, Probability*　　　Wuhan, China

SKILLS

Programming: Python, SQL, C/C++, Perl
Packages: Scikit-learn, TensorFlow, PyTorch, Pandas, XGBoost, Qt, CUDA, NumPy, SciPy, Matplotlib
Tools: AWS, Hadoop/Spark, Git, Docker, UNIX utilities
Modeling: Machine Learning (SVM, Regression, Decision Trees, Clustering), Time Series (ARIMA, HMM, LSTM), Deep Learning (CNN, GCN, RNN, Transformer), Natural Language Processing (word2vec, seq2seq, BERT)

EXPERIENCE

Data Scientist, **Amazon** | Bellevue, WA　　　　　　　　　　　　　　　**May 2021 – Present**
- Hosted meetings with the leaderships from Amazon operational team to understand their business scope, establish project timeline, and negotiate on the final deliveries.
- Performed exploratory data analysis to identify any data quality issues, and feature engineering to pre-process the data.
- Created the multivariate encoder-decoder attention model with two-head outputs to overcomes the conventional time series methods' limitations in long-term forecasting and sparse inputs.
- Setup the data pipeline and deployed the models on native Amazon web service (NAWS) to retrain and make predictions on the weekly basis, saving at least $2MM operational cost per year.

Seismic Imager, **CGG** | Houston, TX　　　　　　　　　　　　　　　**Jan. 2020 –Feb. 2021**
- Provided optimal quality control for the input seismic raw data by analyzing the statistics from billions of seismic records using SQL and mapping out key metrics using Hadoop/Spark big-data tools.
- Processed petabytes of seismic data by detecting anomalies using sparse transformation, removing seismic echoes using adaptive subtraction, and imputing missing features with compressed sensing techniques.
- Helped the client save almost $30MM per production well by providing unbiased interpretation of the geological features using machine learning techniques including SVM and boosted trees.

PROJECTS

Open-Source Software Development for Electron Diffraction Image Processing　　　**Jul. 2018 – Dec. 2019**
- Designed the object-oriented architecture, implemented the software with Python (Github repository: PyRHEED), and provided technical support to users from several research groups across the world.
- Improved the average computational efficiency of the back-end processing module by about 100 times through NumPy vectorization and CUDA parallel computing.
- Built the data pipeline from raw reflection high energy electron diffraction (RHEED) images to processing-ready datasets through autonomous labeling, noise reduction, standardization and featurization.

RHEED Data Analysis With Machine Learning　　　　　　　　　　　　**May 2016 – Dec. 2019**
- Extracted the three-dimensional probability density distribution of the diffracted electron waves from the preprocessed RHEED datasets by learning the parameters of a Gaussian mixture model.
- Simulated thousands of crystal domains based on the Voronoi tessellation using Monte Carlo methods, in order to be combined with the experimentally extracted features for model parameter estimation.
- Estimated the unknown statistics such as the lattice constant, grain size, and preferred orientations from the RHEED images with a Bayesian regression approach.

AWARDS

The Karen & Lester Gerhardt Prize in Science and Engineering at Rensselaer Polytechnic Institute　　**May 2020**
Paul S. Ho '65 Prize in Physics at Rensselaer Polytechnic Institute　　　　　　　　　　**May 2019**
Hillard B. Huntington Award (1976) at Rensselaer Polytechnic Institute　　　　　　　　　**May 2017**
Presidential Graduate Research Fellowship Award at Rensselaer Polytechnic Institute　　　　　**Oct. 2015**