Xi Chen, Ph.D.

San Francisco, CA | 857-209-1002 | billchenxi@gmail.com
stackoverflow.com/story/billchenxi | linkedin.com/in/billchenxi | billchenxi.github.io

SKILLS

Skills:

- Languages: Python, R, SAS, SQL, CUDA, C++, AWS, GCP, Shiny, Heroku, Git, Pytorch, TensorFlow.
- AI/DL/RL/ML Knowledge: https://nvidia.qwiklab.com/public_profiles/5521a192-c2e9-4899-9750-500959646159
- GCP/AWS: https://qwiklabs.com/public_profiles/032b735c-3942-4f65-96e2-46bc821a884a

PROJECTS

SpeedLegal Legal Document Analyzer:

- Building model using machine learning and rule-based approach to analyze legal documents
- Using Python, NLTK, PyTorch, and PDFminer API to build NLP models, used Flask, React, Docker, and QT to build front-end applications.

Deep Learning for Cancer Classification with Gene Expression Data:

- Built a deep learning model to classify cancer types.
- Accomplished a state-of-the-art performance as measured by the accuracy of >97% and the false positive/ negative rates of <0.2% by using transfer learning approach.
- Used Python, TensorFlow, Deep Autoencoder (VAE), Scikit-learn.

Parallelized Interactive Machine Learning on Autonomous Vehicles:

- Used a driving game simulating environment to developed an interactive reinforcement learning model.
- Accomplished a faster model convergence rate as measured by the validation loss over epochs by integrating a human interactive reinforcement learning model.
- Used Python, Unreal Engine API, Convolution Neural Network, Deep Q-learning.

Protein Nuclear Magnetic Resonance (NMR) Reference Correction:

- Built a statistical base model for an estimate of reference correcting values for protein.
- Implemented a Bayesian probabilistic framework to improve the model performance
- Surpassed the state-of-the-art performance as measured by reference error below +/- 0.22 ppm at 90% confidence interval. (State of the art is around 1ppm.)
- Used Python, R, Multi-processing Programming, Statistical Learning, Bayesian, etc.

EDUCATION & TRAINING

University of Kentucky, Ph.D. Bioinformatics & MA Cert. Statistics Aug 2013 to Jun 2019

- **Courses include:** Statistical Analysis, Design & Analysis of Experiments, Computational Inference, Theory of Probability, Intro to Statistical Methods, Regression & Correlation, Statistical Inference, Clinical Trial, Survival and Life Testing, Linear Model & Experimental Design, Longitudinal Data Analysis, Analysis of Categorical Data.
- **Dissertation:** Automatic 13C Chemical Shift Reference Correction of Protein NMR Spectral Data Using Data Mining and Bayesian Statistical Modeling (https://doi.org/10.13023/etd.2019.057)

University of Kentucky, (No-degree) Computer Science Aug 2016 to Jun 2019

- **Courses include:** Machine Learning, Computer Vision, Advanced Data Science, Interactive Machine Learning, Numerical Analysis, Calculus IV, and Linear Algebra.

WORK EXPERIENCE

Verb Surgical 2019 - Present
Data Scientist & Machine Learning Engineer Mountain View, CA

- Worked with a multi-disciplinary team to develop surgical analytics software for a digital surgery platform.
- Improved model performance of more than 5x as measured by accuracy and recall by integrating a video frame data-filtering pipeline and a two-output transform learning model with CNN and LSTM.
- Archived a real-time prediction by integrating the signal process methods.
- Leveraged knowledge in data science, machine learning, statistics, and model scalability.

Technologies: Python, R, Computer Vision, PyTorch, Unit-test, CNN, LSTM, VAE, Docker, etc..

Nvidia Deep Learning Institute 2018 - Present
University Ambassadors / Deep Learning Institute (DLI) Certified Instructor

- To deliver deep learning courses on Computer Vision, Multiple Data Type Analysis, Natural Language Processing, CUDA programming.
- Contents include: Image classification, Object Detection, Image Segmentation, Word Generation, Image and Video Captioning, Text Classification, Text Translation, etc.

Dept. of Statistics, University of Kentucky 2017 - 2019
Research Collaborator Lexington, KY

- Build a High-performance Cluster (HPC) simulation pipeline for the Mix-Gamma Model with R.
- Simulated data from different gamma distributions.
- Implemented unit-test, libraries, and workflow for experiments

Technologies: R, HPC, Slum, Bash, etc..

Dept. of Biochemistry, University of Kentucky 2013 - 2019
Graduate Research Assistant Lexington, KY

- Worked on the construction of Protein NMR Reference Correction and Protein NMR Deuteration Level Detection frameworks.
- Designed and implemented the corresponding software package (BaMORC) and web application.
- Created R package 'BMRBr' for easy download and processing of BMRB files.

Technologies: R, RStudio, Python, Shiny, Docker, etc..

VOLUNTEER EXPERIENCE

Operation Code 2019 - Present
Programmer OperationCode.org

- Serve as a programming volunteer for Operation Code.

Golden Gate Outreach 2019 - Present
Food distributor Golden Gate Park, SF, CA

- Serve as a host to distribute food to the person in need.

PUBLICATION

Papers:

- Finite Mixture-of-Gamma Distributions: Estimation, Inference, and Model-Based Clustering, *Advances in Data Analysis and Classification, May 2019*
- Automatic 13C Chemical Shift Reference Correction for Unassigned Protein NMR Spectra, *Journal of Biomolecular NMR, Aug 2018*
- Parallelized Interactive Machine Learning on Autonomous Vehicles, *NAECON Dec 2018*
- Deep Learning by Doing: The Nvidia Deep Learning Institute, *Journal of Computational Science Education, Dec 2018*
- Pan-Cancer Epigenetic Biomarker Selection from Blood Sample Using SAS®, *MWSUG, Sep 2018*

Packages:

• BaMORC: Bayesian Model Optimized Reference Correction Method for Assigned and Unassigned Protein NMR Spectra, *CRAN, 2018*

• BMRBr: a package that helps R users to analyze data from BMRB data repo by simplifying the download procedure.

Workshops:

• CUDA Programming Workshop, *UK ACM 18*

• Deep Learning for Computer Vision Workshop, *UK ACM 19*

Reward and Certification Highlights

• SIAM International Conference on Data Mining Dissertation Travel Award (2018)

• MWSUG Conference Paper Scholarship (2018)

• Grow with Google Challenge Scholarship (2017/18)

• SC18 Paper Scholarship (2018)

• SAS Certified Clinical Trials Programmer, Statistical Business Analyst, Advanced Programmer

• Deeplearn.ai Certification Series (Coursera)

Preferred Locations: Flexible; currently located in San Francisco, CA
Nationality: Chinese