Sophie Chen

Website Sophie-Chen-Data SophieGarden
217-607-4503 Seattle WA
SophieChenPhD@gmail.com

Sophie Chen enjoys the unique challenges offered by machine learning and data science. She works with large-scale data and applies machine learning and experimentation design toward business problems. She is also a leader for the Seattle chapter of the Women in Machine Learning and Data Science nonprofit. She holds a PhD in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign. She is a creative and insightful problem solver who learns quickly, works effectively, and enjoys working collaboratively with people to solve complex quantitative problems.

Skills

Programming/Tools: Python, Spark, AWS, Sagemaker, TensorFlow, Keras, PyTorch, SQL, Git, Bash, LaTeX, (Basic: Java, Flask, HTML, CSS)

Machine Learning: Supervised Learning, Deep Learning, Recommender Systems, Computer Vision, Unsupervised Learning, Metrics Design, Model Evaluation, Performance Diagnostics

Statistics: Experiments Design, Hypothesis Tests, A/B testing

Certificates: Deep Learning Specialization (DeepLearning.AI), Data Science Nano Degree (Udacity)

Education

PhD, *Electrical and Computer Engineering* *University of Illinois Urbana-Champaign, 2018*

MS, BS, *Optical Science and Engineering* *Fudan University, 2013*

Experience

Applied Scientist - Computer Vision *Zillow, Dec 2021 - current*

- Improved Window, Door, and Open space **object detection** metrics for 3D interactive floor plan composition

Applied Scientist - General Machine Learning *Zillow, April 2019 - Dec 2021*

- Delivered an **embedding similarity search home ranking** model with **deep learning Siamese neural network**, improved similarity search accuracy and efficiency, 90% of actual comparable houses are within recommendation, presented with t-SNE visualization
- Built from scratch a likely user model (xgboost) based on **large scale web click stream data** to **recommend service/product** to potential users, leveraged SHAP for feature selections
- Designed, launched, measured, and presented a **statistical experiment** to shorten the wait time to measure pricing bias and variance by 6 months. From which identified business opportunity and built from scratch a **Huber regresser model** to reduce cost on each home by >$500 (PySpark, Cloud notebook, Python) and presented to >200 stakeholders in all hands
- Deployed models to **production pipelines** that consists of Docker, Airflow, Sagemaker and integrated with client service via GraphQL with minimal engineering support.
- Built reporting systems to **monitor model performance** in terms of production stability as well as business metrics (DataDog, Luminaire, Mode)

- Diagnosed, fixed, and reported an **A/B testing experiment** with covariates control to pick the best offer route that max out profitability and user experience
- Led a data analysis that enabled **offer automation expansion** by >50%
- Mentored juniors and interns. Led Women in Tech and other affinity network events

Data Scientist Fellow *Insight Data Science, 2018*

- Developed a Chrome Extension using **customized hierarchical classification** algorithm with natural language processing (**NLP**) to predict best folder for downloading files
- Achieved 82% predictions on the right path, of which 69% within 1 level away
- Deployed with Python, Flask, Chrome API, PyInstaller (**strong programming skills**)

Optical Modeling Researcher *University of Illinois, Fudan, 2010-2018*

- Build **predictive** thermal model with **3D visualization**, improved fiber laser power by 10%+, and broadband fiber source power to **the highest to date** (Matlab)
- Verified three-body formation theory for **the first time** via automated data collection and **applied statistical modeling in physics**
- Analyzed **transformed** spectra and **time series data** with computational models, reduced RMSE of molecular potential curves by 20%+, discovered 11+ new electronic transitions
- Developed **playbook-style documentation** for managing various lab tasks

Projects

Various Computer Vision projects

- Build a **ResNet** architecture from scratch using **TensorFlow, Numpy**
- Designed **neural network** image classifiers, image segmentation, and object detection, face recognition, art style generation with **PyTorch, TensorFlow** (CNN, ResNet, VGG, YOLO, U-Net)
- Built a Command Line App to automate extract, load, transform (**ETL**), learning, and inference

RNN sequence models, Transformers

- Build **RNN** sequence models from scratch using TensorFlow, Numpy
- Built **pipeline** several NLP projects using **Word Embeddings, Sequence Models, Attention Mechanism, and Transformers**

Renter Eviction Consulting

- Analyzed racial bias in renter evictions with **propensity modeling**, reported with **interactive visualization** (Tableau)
- Achieved 0.7 RMSE on predictions of eviction rate percentages, required extensive **cleaning and engineering** of high-dimensional data (XGBoost, Cross Validation)