Xinru Liu (She/her/hers) — Resume

+15083691863 • liu.xinruuu@gmail.com • xliu-522 • xinru-liu-64865a123

Education

Boston University (*Ph.D. Candidate in Statistics*) **Sept 2019 - Dec 2024 (expected)**

- **Related coursework**: Machine Learning, Reinforcement Learning, Generalized Linear Model & Application, Computational Statistics, Stochastic Processes, Time series, Bayesian Statistics, Hypothesis Test, Prob Theory.
- **Qualifying Exam**: Applied Statistics, Mathematical Statistics

Wheaton College (*B.A. in Mathematics & Computer Science, minor in piano performance*) **Sept 2015 - May 2019**

Summa Cum Laude, GPA: 3.93 *Norton, MA*

- Honors: Phi Beta Kappa Ruth Redding Graduate Scholarship (2019), Helen Zoe Duncan Prize in Piano Performance (2017-2019).
- **Related coursework:** Data Structure, Algorithm, Theory of Computing, Graphics, Data Visualization, Database Management.

Technical & Programming Skills

Programming: Python, R, Matlab, C & C++, SQL

DL: PyTorch, Keras, HuggingFace, Transformers, Cuda

Visualization & EDA: Numpy, Pandas, Matplotlib, Seaborn, Jupyter, Weights & Biases, ggplot, dplyr

Operating Systems: Linux, Windows, Mac OS

ML Frameworks: SKLearn, SciPy, XGBoost

Tools/platforms: Git, AWS (Cortex), Microsoft Azure ML, Jira, Confluence, LabelBox

Work Experience

Applied Science Intern **June 2023 - August 2023**

Liberty Mutual Insurance Solaria Labs *Boston, MA*

- **Background:** Enhanced underwriting decisions by fine-tuning LLM using **Human In The Loop** approach.
- Improved the **Flan-T5 LLMs'** performance in classifying risk characteristic queries (**10% ↑ in F1-score**), using **prompt engineering** and the **Reinforcement Learning from Human Feedback** fine-tuning algorithm(**PPO**).
- Reduced the GPU memory requirement and computational cost by integrating **LoRA** reparameterization technique.
- Facilitated reproducibility among team members by maintaining a **GitHub repo** adhering to industry best practices.

Graduate Data Science Intern **June 2022 - August 2022**

Liberty Mutual Insurance Office of Data Science *Boston, MA*

- **Background:** Identified similar business descriptions across diverse economic sectors using Hierarchy in NAICS Code.
- **Approach 1:** Improved the baseline accuracy of semantic search for code prediction by **4%** by developing a customized hierarchical similarity metric on **MPNet** embeddings through Breadth-first search (**BFS**) algorithm.
- **Approach 2:** Boosted classification accuracy **from 87% to 95%** by engineering a **hierarchical NLP classifier** that constructs a layered representation linking business embeddings to a structured hierarchy of NAICS codes.

Research & Projects

A Statistical Perspective on Algorithm Unrolling Method for Inverse Problems **Sept 2021 - Jan 2023**

- Derived a theoretical error rate for estimating nonparametric regression models using a sparse Bayesian **gradient descent network (GDN)**, an algorithm unrolling deep neural network architecture driven by proximal gradient descent.
- Addressed **large-scale** inverse problems using the GDN in **image reconstruction** by denoising blurred images.
- Resolved the computational intractability issue in sparse deep learning by leveraging an efficient approximate **MCMC algorithm** (SA-SGLD) that employs asynchronicity and sparsity for fast sampling (**Pytorch**).

Thompson Sampling in Dynamic Pricing **Sept 2022 - present**

- Maximize cumulative revenue by developing and implementing a **Bayesian dynamic pricing** policy within a recommender system to optimize product pricing over time.
- Efficiently learn the unknown noise CDF of the pricing model by training a derivative-based neural network.
- Employ **Thompson Sampling**, a **Bayesian contextual bandit** posterior sampling algorithm, to efficiently estimate model parameters and derive posterior contraction rates.

Reinforcement Learning through soft Deep Q-Network **Jan 2021 - present**

- Extend the foundational Q-learning method by pioneering a **soft Deep Q-Network**, where an agent learns effective policies directly from **high-dimensional** inputs while delivering precise actions within **continuous** action spaces.
- Establish the algorithmic and statistical error rates for approximating the ground true action-value functions using the iterative policy sequence achieved through soft-DQN.

Publications

- **Liu X., Atchade Y.** Neural Thompson Sampling for Dynamic Pricing. *Unpublished Manuscript*.
- **Wang L., Liu X., Matthew A., Atchade Y.** On cyclical MCMC sampling. *Accepted by AISTATS 2024*.
- **Atchade Y., Liu X., Zhu Q.** A statistical perspective on unrolling models for inverse problems. arXiv:2311.06395. *Under review by JMLR*.
- **Liu X., Gui X., Qi T., Guo W.** (2018). Multimodal Data Fusion in 3D printing Quality. *IEEE Sensors Letters*, DOI 10.1109/LSENS.2018.2881475.