

YANGHONG GUO

2 years working experience and 10 years math/statistics background

☎ 682-298-2152 ✉ yxg190031@utdallas.edu 🏠 Website 🔗 LinkedIn 📍 Dallas, TX

EDUCATION

University of Texas at Dallas

Ph.D. in Statistics; Bayesian/ML Methods in Biostatistics; GPA 3.97/4

Aug 2021 – May 2025 (Expected)

Dallas, TX

Columbia University

M.A. in Statistics

Sep 2016 – May 2018

New York, NY

East China University of Science & Technology

B.S. in Mathematics & Applied Mathematics

Sep 2012 – Jun 2016

Shanghai, China

EXPERIENCE

Research Assistant (PhD)

UT Dallas

Jan 2023 – Present

Dallas, TX

Industry Researcher (Full-time)

Bank of China

Jul 2019 – Aug 2020

Beijing, China

- Analyzed bank customer service data by the NLP Semantic system
- Detected and predicted frequently happened customer issues

Bitcoin and Crypto Data Analyst (Full-time)

Beijing Micai Investment Co., Ltd.

Jul 2018 – Mar 2019

Beijing, China

- Implemented automatic quantitative trend-tracing trading strategy with Support Vector Regression (SVR)
- Applied web-crawler with Python to collect online STO data then saved by MYSQL

Quantitative Analysis Intern

China Merchants Securities Co., Ltd.

Jun – Aug 2017

Shenzhen, China

PROJECTS

Bayesian Integration for Bulk RNA-Seq Data and Spatial Transcriptomics | [GitHub](#) | *R, C++*

Jun 2023

- Applying Bayesian method for spatial domain identification of spatial transcriptomics (ST), achieving accuracy of 90%
- Leveraging the identified domains to deconvolute Bulk RNA-Seq data to spatial-domain level
- Integrating clinical metadata with spatially-deconvoluted Bulk RNA-Seq insights

Bayesian DM Model for Integration of Clinical and Single Cell Data | [GitHub](#) | *R, C++*

Sep 2022

- Proposed a hierarchical Bayesian framework for the integration of single-cell data and covariates
- Implemented the model on a self-identified tree structure of cell types
- Discovered the relationship between the cell type abundance and covariates that were not studied before

Noise-Robust Learning for Distantly-Supervised Entity-Relation Extraction | [GitHub](#) | *PyTorch*

Jun 2022

- Incorporated a pre-trained transformer into sequence tagging scheme for distantly-supervised joint extraction
- Proposed a bootstrap learning framework with a noise-robust loss to select high-quality instances dynamically

Change-point Detection Using Bayesian Inference | [GitHub](#) | *R, MCMC*

Mar 2022

- Applied t-shrinkage prior and Horseshoe before historical DJI return and Bitcoin-USD return during fluctuating time
- Implemented MCMC methods with R and visualized the detected change points
- Detected 100% change-points successfully by the given criteria to define a change-point

Potential ETC Customer Identification | *Python, TensorFlow, Scikit-learn*

Nov 2019

- Processed and cleaned the original customer dataset of 9 million samples with de-noise analysis
- Utilized data discretization methods to further reduce batch size and model complexity
- Refined the data by Grid Search and Ensemble Generation and achieved an AUC around 0.85 when targeting a potential ETC Customer (rank 10/200)

SKILLS

- **Languages:** Python, MATLAB, C++, LaTeX, HTML
- **Data Analysis:** R, SAS, MySQL, SQL Server
- **Version Control:** GitHub, Git, Google Colab
- **Machine Learning:** TensorFlow, Scikit-learn, PyTorch