Tong Si

tong.si@slu.edu | (314)-696-9062 | St. Louis, MO
Homepage: https://mathstat.slu.edu/~gong/tong.html

Educational Background

Saint Louis University St. Louis, MO

- **Ph.D. Candidate** in **Statistics**(GPA 3.97/4.0) *Aug. 2020-Present*
  Advisor: Dr. Haijun Gong
  Department of Mathematics and Statistics
- **M.S. Candidate** in **Artificial Intelligence** (GPA 3.97/4.0) *Aug. 2022-Present*
  Department of Computer Science
- **M. A**. in **Mathematics** *Aug. 2020 - May 2022*
  Department of Mathematics and Statistics

Jilin University Changchun,China

- **B. S.** in Mathematics and Applied Mathematics. *Sept. 2016 - Jun. 2020*

Research Publications

Peer-Reviewed Paper

1. (under review) Tong Si *, Yunge Wang, Lingling Zhang, Kate Cannell, Tae-Hyuk Ahn, Haijun Gong * "Multivariate time series change-point detection with a novel Pearson-like scaled Bregman divergence"
2. **Tong Si**, Zackary Hopkins, John Yanev, Jie Hou, and Haijun Gong. "A novel f-divergence based generative adversarial imputation method for scRNA-seq data analysis." *Plos ONE* 18, no. 11 (2023): e0292792.
3. Richards, Helen, Yunge Wang, **Tong Si**, Hao Zhang, and Haijun Gong. "Intelligent Learning and Verification of Biological Networks." *Advances in Artificial Intelligence, Computation, and Data Science: For Medicine and Life Science* (2021): 3-28.

Peer Reviewed Abstract

4. **Tong Si**, Yunge Wang, Lingling Zhang, Kate Cannell, and Haijun Gong. " Change-Point Detection Using Scaled Bregman Divergence." *F1000Research, 22nd International Conference on Bioinformatics (2023)*
5. **Tong Si**, Zackary Hopkins, John Yanev, Jie Hou, and Haijun Gong. " sc-fGAIN: An f-divergence-based Generative Adversarial Imputation Method for scRNA-sq Data Analysis." *F1000Research, 22nd International Conference on Bioinformatics (2023)*

Research Projects

Imputation of Time Series Data via Generative Models and GRU *Oct. 2023-Present*
Team leader, Dr. Gong's group

- Conduct a thorough literature survey on time series data imputation to identify prevalent limitations and gaps in current methodologies.
- Develop a GRU-based time-series generative adversarial imputation network algorithm and investigate the mathematical theory underlying the algorithm.
- Implement the time series imputation algorithm based on different divergence functions using Python.
- Prepare a manuscript for submission to a peer-reviewed journal.

Change-Point Detection for Time Series Data Using Scaled Bregman Divergence *June 2023 - March 2024*
Team leader, Dr. Gong's group

- Developed a Pearson-like Scaled Bregman Divergence Method [3] for Change-point Detection (CPD) of multivariate time series data
- Investigated the mathematical foundation of the algorithm and reinforce the algorithm's generality and reliability across a broader range of applications.
- Reproduced comparative methods in R and Python to benchmark our model against existing techniques. Compare the accuracy in identifying change-points, and performance across diverse datasets and conditions.
- A paper is under review in a peer-reviewed journal.

Innovative Web-Based Library Management System *Sept. 2023 - Dec. 2023*
Team leader of Course Project

- Utilized SQL for robust database design and management, ensuring efficient data storage, retrieval, and manipulation; Implement the user interface using HTML, creating an intuitive and responsive web application.

- Built the core functionality of the system using Python, ensuring seamless integration with the database and frontend components; Apply GitHub for source code management and team collaboration, maintaining an organized and efficient development workflow.
- Used CircleCI for continuous integration, automating code testing and deployment processes, to enhance code quality and deployment efficiency
- Employed Docker Hub for containerizing the application, ensuring consistent deployment across different environments.

Imputation of sc-RNA Sequencing Data via Generative Adversarial Networks *Oct. 2022 -May. 2023*

Team leader, Dr. Gong's group

- Led the team to develop a novel single cell f-divergence based generative adversarial imputation network (sc-fGAIN) algorithm to impute the missing values in the single cell RNA sequencing data.
- Implemented the sc-fGAIN algorithm using Python and provide mathematical proofs to confirm its effectiveness and general applicability in imputation tasks.
- Managed a massive dataset with dimensions 10,164 by 3,918, ensuring efficient data preprocessing and algorithm application.
- Implemented and compared different state-of-the-art imputation methods as benchmarks using R, Python, and MATLAB to validate the superiority of our approach.
- Our paper [1] has been **published** in PLOS ONE in 2023, and receive a **Best Oral Presentation Award** [4] at 2023 International Conference on Bioinformatics, held in Brisbane, Australia.

Analytical Text Processing Using Machine Learning *Sept. 2022 - Nov. 2022*

Course Project

- Applied Python libraries Pandas for data manipulation and Scikit-Learn for machine learning model implementation, including using feature sklearn.feature_extraction.text.CountVectorizer for text preprocessing and feature extraction
- Processed raw text data using tokenization and lemmatization techniques.
- Implemented a variety of classification algorithms, including Naive Bayes, SVM, and Random Forest, to compare performance. Optimize models using cross-validation and grid search techniques.

Statistical Inference and verification of Regulatory Networks *Sept. 2020 - May 2021*

Collaborative Research Project, Dr. Gong's group

- Applied a weighted dynamic Bayesian network method to reconstruct gene regulatory network from time series microarray data with other team members.
- Implemented different model checking technique, including SMV and PRISM for the network verification.
- Wrote a manuscript in collaboration with other team members and **a paper [2] was published in 2021**.

Teaching and Professional Service

Research Assistant, Dr. Gong's Group, Saint Louis University *Jan. 2023-Present*

Reviewer for the following Journals: *2021- present*

BMC Bioinformatics; Journal of Bioinformatics and Computational Biology; PLOS ONE; Heliyon; Journal of Theoretical Biology; Genomics …

Teaching Assistant for Bayesian Statistics *Aug.2023- Dec. 2023*

Treasurer of Association for Women in Mathematics (AWM) in St. Louis University *Aug. 2022- Jan. 2023*

Instructor of College Algebra, Saint Louis University *Jan. 2022-Dec. 2022*

Teaching Assistant of Calculus I, Saint Louis University *Aug. 2021 – Dec. 2021*

Conference Presentation

Oral Presentation **at the Mathematical Association of America Missouri Section, Liberty, MO** Apr. 2024

Oral Presentation **at the Danforth Plant Sciences Center, St. Louis, MO** Jan. 2024

Oral Presentation, **22nd International Conference on Bioinformatics, Brisbane, Australia** *Nov. 2023*

Poster Presentation, **International Conference on Intelligent Biology & Medicine, Tampa, FL** *Jul. 2023*

Awards and Certificate

GLBIO 2024 travel fellowship *Mar.2024*

Dean's Travel Award, Saint Louis University *Nov. 2023, Mar.2024*

Best Oral Presentation Award, 22nd International Conference on Bioinformatics, Australia *Nov. 2023*

Travel Award, Forty Third Midwest Probability Colloquium *Oct. 2022*

Teaching Certificate, Saint Louis University *Sept. 2024*

TECHNICAL SKILLS AND CERTIFICATIONS

- Computer Languages: Python, R, MATLAB, SQL, HTML
- Skills: Data analysis for big data, Software development, Database skills, Website building skills