Qimin Zhang

Email: qqz5133@psu.edu Phone: +1(814)529-3984
Website: https://qiminzhang77.github.io/

Education

- **The Pennsylvania State University** State College, USA
  Ph.D in Computer Science and Engineering *Aug. 2019 – Jun. 2023(expected)*
- **University of Chinese Academy of Sciences** Beijing, China
  Master of Engineering in Computer Technology *Sep. 2016 – June. 2019*
- **Beihang University** Beijing, China
  Bachelor of Engineering in Aircraft Airworthness *Sep. 2012 – June. 2016*
  Bachelor of Science in Applied Mathematics (minor)

Skills

- **Languages** C++, C, Python, Shell
- **Technical Skills and Tools** Machine Learning (scikit-learn, PyTorch), Bioinformatics (SAMtools, STAR, HISAT2, bedtools, etc.), Cloud Computing (AWS), Web Development (HTML/CSS/JS), Linux, Git, Docker

Work Experience

- **Laboratory Corporation of America Holdings (LabCorp)** Remote
  Data Science Summer Intern *May 2022 - Aug 2022*

  Predict gut metabolites from gut microbiome data using machine learning
  - Developed a set of highly accurate data pipelines to predict gut metabolites and metagenome functions from gut microbiome data
  - Managed to work on a Colorectal Cancer dataset and achieved all metabolites well predicted (using Spearman correlation coefficient as the metric)

  AWS resource access control
  - Developed a web application to enable the AWS resource access control and deployed to ECS
  - Developed on service end to create IAM roles to access AWS resource (S3 and DynamoDB) and developed on front end to implement the user interface

Research Experience

- **Bioinformatics & Applied Machine Learning** Sep 2016 - Present
  - **Transcriptome assembly & single-cell RNA-seq data analysis** Developed Scallop2, a transcriptome assembler that enables accurate assembly at both single-cell resolution and bulk level. Designed and implemented a dynamic programming algorithm and an enhanced consensus algorithm to improve 85.9% and 46.6% in precision comparing with two leading tools at the same level of sensitivity.
  - **High-throughput computing** Designed a density-based clustering model for resource allocation problem in high throughput computing. Managed to work on four bioinformatics workflows and achieved over 49% memory saving compared with fixed resource allocation strategy.
  - **Healthcare** Explored machine learning techniques to biomedical problems. Extracted feature of sputum sound signals using wavelet transform algorithm. Implemented a BPNN model and improved the precision of sputum sound detection to 84.53%.

Selected Publications

1. **Qimin Zhang**, Qian Shi, Mingfu Shao. Accurate assembly of multi-end RNA-seq data with Scallop2. *Nature Computational Science*, 2, 148-152, 2022.
2. **Qimin Zhang**, Nathaniel Kremer-Herman, Benjamin Tovar, Douglas Thain. Reduction of workflow resource consumption using a density-based clustering model. *2018 IEEE/ACM Workflows in Support of Large-Scale Science (WORKS)*, pages. 1-9, 2018.