

Tengfei Cui

+1 (206) 209 9522) - [Personal Website](#) [Github](#) [Email](#)

EDUCATION

University of Washington, Seattle

MS Biostatistics (Thesis Track)

Seattle, USA

09/2023 - 06/2025

Xi'an Jiaotong-liverpool University

BSc Applied Mathematics

Suzhou, China

09/2019 - 06/23

RESEARCH EXPERIENCE

Differential Allelic Expression Using Single-cell Data

Research Assistant, University of Washington, Advisor: Prof.G.Qi

10/2023 - Present

- Building a new model to study the gene expression by single-cell data. Add a prior distribution to current model by using information across different genes to improve the performance of the model.
- Accelerate to fitting data with the help of GPU. Reduce the training time to get accurate estimates by parallel computing and distributed computing.

Survival Analysis on MIMIC III Database

Research Assistant, Xi'an Jiaotong-Liverpool University, Advisor: Prof.X.J.Zhu

5/2022 - 10/2022

- Building survival analysis models to predict survival time of patients with malignant neoplasm of liver.
- Extracted and processed lab test data of patients with malignant neoplasm of liver and get an insight of dataset by drawing K-M curves
- Built Cox PH models with a single predictor, chose potential risk factors, and built Cox PH models with multiple predictors
- Improved performance of Cox PH model by best subset selection, AIC value, and linear regression models

Automated Clinical Coding By Deep Learning

Research Assistant, University of Hong Kong, Advisor: Prof.L.Yu

5/2022 - 10/2022

- Building a new automated medical code prediction system which could transform discharge summaries of the MIMIC III dataset into ICD-9 codes.
- Finetuned several outstanding pre-trained language models, including BERT, BioBERT, ClinicalBERT and BlueBERT, in multilabel text classification tasks and compared the performance with the CNN-based models
- Tried new long sequence transformer models, such as Clinical-Longformer and Clinical-BigBird, to improve the performance of transformer-based models
- Designed new hierarchical fine-tuning architectures to extend the maximum input sequence of transformer-based models

Impact of COVID-19 on Academic Research – A Data-Driven Study

Research Assistant, Nanjing University & University of Virginia, Advisor: Prof.Q.Du

12/2021 - 11/2022

- Collectively leveraged an array of machine learning techniques to study the impact of the ongoing COVID-19 pandemic on the number and quality of the recently published academic papers.
- Leveraged TFIDF numerical statistics for keywords extraction and latent semantic analysis (LSA); clustered semantic vectors using cosine similarity, allowing summarization of semantic themes
- Utilized regular expression, n-gram model, and self-attention mechanism to drive pattern extraction, with a focus on converting semi-structured resume data into fully structured data
- Facilitated knowledge graph construction using Neo4j graph database, with a focus on subject-predicate-object three-element groups

ACTIVITIES

- **Group Reading** in University of Washington with Prof. R.Guo - 02.2024 - Present
- **Group Reading** in University of Washington with Prof. G.Chen - 03.2024 - Present

TECHNICAL SKILLS

Programming Languages: R, Python, C++