Shichao Feng

E-mail: uranusfeng1992@gmail.com

EDUCATION

Ph.D. of Computer Science and Engineering in University of North Texas

expected Aug 2025

- Direction: bioinformatics, biocomputing, deep learning, machine learning, optimization, big data
- Supervisor: Dr. Xuan Guo

M.S of Computer Information and Technology in Purdue University

2016 - 2018

- Direction: deep learning, data mining, NLP, health surveillance, social media
- Supervisor: Dr. Keyuan Jiang

B.S of Software Engineering in Donghua University

2011 - 2015

- Scholarship for Academic Excellence

RESEARCH AND PROJECT

Comprehensive tools of probiotic protein identification based on data-driven deep learning frameworks - University of North Texas, supported by NIH

2022 - present

- Retrieved all up-to-date human gut metaproteomics studies and results; processed and extracted necessary data (matches between peptides and mass spectrum) into tabular format.
- Applied multi-layer statistical filters to select high-quality data for training models.
- Researched variants of CNN and Transformer-based deep learning architectures and different training strategies to develop a high-performance deep learning model.
- Implemented a deep-learning-based search engine for peptide and protein identification by integrating models with pre/post-processing scripts via C++ distributions of PyTorch (LibTorch).
- Optimized the inference speed of peptide-spectrum matches by paralleling the similarity calculation process using OpenMP.
- The model and post-processing codes are available: https://github.com/Biocomputing-Research-Group/WinnowNet

Protein inference in metaproteomics using mathematical optimization methods integrated with metagenomics (DNA) - University of North Texas, supported by NIH

2020 - 2022

- Paralleled multi-omics (DNA and protein) analysis with more than 100GB of raw data from whole genome sequencing and mass spectra on high-performance computing systems
- Researched mathematical optimization methods to develop a novel solver incorporating multiple omics information aiming at handling protein inference problem
- Conducted the end-to-end protein analysis using the solver for microbiomes in the ocean, soil and human gut to explore new biological findings
- The solver is provided as a tool: https://github.com/Biocomputing-Research-Group/MetaLP

Peptide identification in metaproteomics based on deep learning neural network -- University of North Texas, supported by NIH

2018 - 2020

- Constructed datasets from 20GB shotgun metaproteomics mass spectra data using high performance computing
- Utilized a convolutional neural network to train a peptide identification model for similarity calculation of peptide-spectrum matches.
- Packaged the model as a tool for inference: https://github.com/Biocomputing-Research-Group/DeepFilter

Shichao Feng

Smm4hSharedTask-- University of North Texas, data challenge

2020

- Preprocessed tweets data and implemented classification model for extremely imbalanced tweets including multilingual tweets that mention medication
- Applied multilingual models to augment data for rare non-English tweets and SMOTE for data imbalance
- Conducted experiments with multiple deep learning (machine learning) methods including Bi-LSTM, attention mechanisms, and logistic regression with engineered features: https://github.com/ShichaoFeng92/Smm4hSharedTask

Detection of Potential Drug Effect from Twitter – Purdue University Northwest, supported by NIH

2016-2018

- Used convolutional and LSTM neural network to identify Personal Experience Tweets
- Crawled medicine-related Twitter posts and construct corpus of 55M tweets in specific data domain
- Conducted statistical analysis of identified personal experience tweets from social media to select the most significant tweets with people's experience on medication
- Analyzed the most popular medicines and mine the potential effects of them with relational similarity

Decision Support System Based on Business Intelligence – Donghua University

2015-2016

- Constructed the data warehouse to decrease the response time of the Decision Support System
- Used the Integration Services of Microsoft SQL Server for the ETL processes

Virtual 3D Traversal Website Based on VRML – Donghua University, recommended project of the College Students' Innovative Entrepreneurial Training in Shanghai

2012-2015

- Used Maya to build 3D model and embedded scripting language
- Improved professional skills and deepened my understanding of WebGL

Parallel Programming and Calculating Course Website with SSH Framework – Donghua University, Bachelor thesis

2015

- Designed UI interface and database and wrote programs of the website with script
- Comprehended MVC model website with SSH framework

PUBLICATION

- *Proteomic stable isotope probing with an upgraded Sipros algorithm for improved identification and quantification of isotopically labeled proteins.* With Yi Xiong, Ryan S. Mueller, Xuan Guo and Chongle Pan, Microbiome (Accepted)
- *SEMQuant: Extending Sipros-Ensemble with Match-Between-Runs for comprehensive quantitative metaproteomics.* with Bailu Zhang, Manushi Parajuli, Yi Xiong and Xuan Guo, International Symposium on Bioinformatics Research and Applications (ISBRA). Singapore: Springer Nature Singapore (Accepted).
- *CloudProteoAnalyzer: scalable processing of big data from proteomics using cloud computing.* with Li, Jiancheng, Yi Xiong, Chongle Pan, and Xuan Guo, Bioinformatics Advances, 2024
- *FineFDR: Fine-grained Taxonomy-specific False Discovery Rates Control in Metaproteomics.* with Shengze Wang, Chongle Pan, and Xuan Guo, IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2022
- *MetaLP: an integrative linear programming method for protein inference in metaproteomics.* with Hong-Long Ji, Huan Wang, Bailu Zhang, Chongle Pan and Xuan Guo, PLOS Computational Biology, 2022
- *Deep learning for peptide identification from metaproteomics datasets.* With Ryan Sterzenbach, and Xuan Gu, Journal of Proteomics, 2021
- *Identifying tweets of personal health experience through word embedding and LSTM neural network.* with Keyuan Jiang, Qunhao Song, Ricardo A. Calix, Mtrika Gupta, Gordon R. Bernard, BMC bioinformatics, 2018.

Shichao Feng

- *Detecting Personal Experience Tweets for Health Surveillance Using Unsupervised Feature Learning and Recurrent Neural Networks*, with Keyuan Jiang, Jiyun Li, Ricardo A. Calix, Mtrika Gupta, the 32nd AAAI workshop on Health Intelligence, 2018.
- *Assessment of word embedding techniques for identification of personal experience tweets pertaining to medication uses*, with Keyuan Jiang, Ricardo A. Calix, Gordon R. Bernard, Precision Health and Medicine: A Digital Revolution in Healthcare, 2019.
- *Mining Potential Effects of HUMIRA in Twitter Posts through Relational Similarity*, with Liyuan Huang, Tingyu Chen, Gordon R. Bernard, Medical informatics Europe, 2020

CONFERENCE

The 20th International Symposium on Bioinformatics Research and Applications (ISBRA), 2024. Attending (upcoming)
The 32nd AAAI workshop on Health Intelligence, 2018. Oral presentation.

WORKING EXPERIENCE

Research Assistant/Teaching Assistant in University of North Texas	**2018-now**
Research Assistant/Lab Assistant in Purdue University and Purdue University Northwest	**2016-2018**
Research Assistant/Teaching Assistant in Donghua University	**2015-2016**
Data warehouse construction and Java Web application in Hesheng Lab in China	**2015-2016**

ACADEMIC SERVICE

Reviewer, Journal article, Health Information Science and Systems	**2023-2023**
Reviewer, Journal article, PLOS Computational Biology	**2022-2023**
Reviewer, conference paper, IEEE International Conference on Bioinformatics and Biomedicine (BIBM)	**2022-2023**
Reviewer, conference paper, International Journal of Bioinformatics Research and Applications	**2022-2022**
Reviewer, Journal article, PLOS One	**2022-2022**
Reviewer, conference paper, IEEE International Conference on Smart City Innovations	**2021-2021**

EXTRA CURRICULAR EXPERIENCE

Patent “Foam glass ceramic and preparation method thereof” (CN102060444B)	**2010-2030**
Patent “Industrial wastewater treatment device” (CN201942575U)	**2010-2020**
Volunteer in Shanghai Science & Technology Museum China	**2011-2015**
National second-level Taekwondo athlete, China	**2008-2014**

SKILLS

TECHLOGICAL SKILLS: python, java, C/C++, Unix-Shell, HPC, SQL, Matlab