

Chengshuai Zhao

☎ +1(949)227-7633 | ✉ chengsz4@uci.edu | 🌐 chengshuaizhao0.github.io | 📍 Irvine, CA, USA

EDUCATION

University of California - Irvine Sep 2022 - present

- Exchange Program of Data Science

Huazhong Agricultural University Sep 2018 - present

- B.S. in Data Science, GPA: 3.97/4.0, Ranking: 1/63

- B.S. in Botany, GPA: 3.87/4.0, Ranking: 2/364

PUBLICATIONS & COPYRIGHTS

- [1] J. Xia†, **C. Zhao†**, B. Hu, Z. Gao, C. Tan, Y. Liu, S. Li, and S. Z. Li, “Mole-bert: Rethinking pre-training graph neural networks for molecules,” accepted as a conference paper at ICLR, 2023.
- [2] **C. Zhao**, H. Wang, W. Qi, and S. Liu, “Toward drug-mirna resistance association prediction by positional encoding graph neural network and multi-channel neural network,” *Methods*, vol. 207, pp. 81–89, 2022.
- [3] X. Xu, S. Liu, Z. Yang, X. Zhao, Y. Deng, G. Zhang, J. Pang, **C. Zhao**, and W. Zhang, “A systematic review of computational methods for predicting long noncoding rnas,” *Briefings in Functional Genomics*, vol. 20, no. 3, pp. 162–173, 2021.
- [4] **C. Zhao**, S. Liu, F. Huang, S. Liu, and W. Zhang, “Csgnn: Contrastive self-supervised graph neural network for molecular interaction prediction,” in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, Online*, 2021, pp. 19–27.
- [5] **C. Zhao**, Y. Zhou, J. Cheng, and S. Liu, *Drug Interaction Data Collection and Analysis System*, V1.0. Software Copyright, National Copyright Administration, P.R.C., 2021, 2021SR1429280.
- [6] **C. Zhao**, Y. Qiu, S. Zhou, S. Liu, W. Zhang, and Y. Niu, “Graph embedding ensemble methods based on the heterogeneous network for lncrna-mirna interaction prediction,” *BMC genomics*, vol. 21, no. 13, pp. 1–12, 2020.

RESEARCH EXPERIENCE

Research Intern, Center for Artificial Intelligence Research and Innovation, Hangzhou, China

Pre-training Strategies for Graph Neural Networks Jul 2022 - present

- Discovered a potential reason for the negative transfer issue of node attribute prediction in graph classification pre-training task can be attributed to extremely small and unbalanced vocabulary
- Established a query system that searches and visualizes most similar graphs with input by computing and integrating similarity derived from model-developed graph embedding and inherent graph structure

Undergraduate Researcher, BioMedical Big Data Mining Lab, Wuhan, China

Drug-miRNA Association Discovery May 2022 - Sep 2022

- Proposed a computational method, which updates the node and positional features by separate channels

while keeping permutation and rotation equivariance, and synthesize from various operator-based perspectives via a positional encoding graph neural network and multi-channel neural network, to enable drug-miRNA association and prompt drug discovery process

Link Prediction

Aug 2020 - Aug 2021

- Collected, processed, and constructed homogeneous and heterogeneous networks with 40,00+ nodes and 500,00+ edges
- Proposed a novel graph neural network (GNN) framework based on self-supervised learning and contrastive learning, which generated pseudo-labeled data to facilitate representation learning from the limited samples
- Implemented the algorithm with PyTorch, achieved a maximum 20% improvement in AUC values and 35% improvement in AUPR values compared to state-of-the-art methods (e.g., node2vec, Line, and GAT) on 7 publicly available datasets
- Analyzed experimental results and pipelined the framework with the t-SNE algorithm, Matplotlib and MS Visio tools
- Summarized the ideas as a manuscript and published it at a top-tier conference

LncRNA-MiRNA Interaction Inference

Feb 2020 - Dec 2020

- Developed a high-accuracy algorithm entailing ensemble learning and attention mechanism based on the heterogeneous network to predict lncRNA-miRNA interactions and prompt to reveal potential genetic disease
- Architected the model with Scikit-learn and TensorFlow framework, conducted 5-fold cross-validation (CV) experiments, and advanced the predictive capacity (5.12%-562.84% improvement in F1 score)
- Design and conducted ablation study and case study to validate proposed methods both theoretically and practically

Long Noncoding RNAs Classification

Oct 2019 - May 2020

- Reviewed previous studies, and developed a comprehensive Python package ezLncPred which integrated 9 machine learning based models and a convenient command-line method to assist researchers who intended to identify lncRNAs

COURSE PROJECTS

Machine Learning Algorithm

Oct 2021 - Dec 2021

- Implemented Linear Regression, Logistic Regression, Decision Tree, Naive Bayes, Principal Component Analysis, and K-means Clustering algorithms with Python package NumPy and SciPy

Drug Analysis System

Jul 2021 - Aug 2021

- Designed and built a lightweight Python-based drug analysis system that allowed drug interactions real-time search, online acquisition, group-based statistics, network visualization, and similarity-based filter to promote drug development
- Utilized Selenium based multi-threading web crawler to extract drug structure, attribution, and interaction information from dynamic web pages, collected and processed 1,350,000+ drug interactions with Xpath location technique
- Implemented backend API endpoints with PyMongo and pass on queries to the MongoDB database, developed a friendly frontend Graph User Interface and implemented with TkInter and PyQt5
- Modeled a comprehensive algorithm that combined Jaccard similarity coefficient based structure and

node2vec embedding base topology to filter candidate drugs and predict potential interactions via multi-view similarity

Gene Expression Mutation Detection

Jul 2021

- Designed and architected Markov chain Monte Carlo (MCMC) based methods with PyMC3 to precisely detect the mutation point of ENSG00000000419 gene expression in the human body

ACTIVITIES

Chief Director, Huazhong Agricultural University, Wuhan, China

Yunshu Club Academic Salon

Aug 2020 - Aug 2021

- Led and managed an academic organization including 20 undergraduate students, where weekly academic seminars and regular salons are conducted to brainstorm and share literature and projects

HONOURS & AWARDS

Honorable Mention, Mathematical Contest in Modeling

Apr 2021

- Awarded by: Consortium for Mathematics and Its Application (COMAP)

Bronze Prize, Mechanisms of Action (MoA) Prediction

Dec 2020

- Awarded by: Kaggle

Third Prize, 13rd Chinese College Students Computer Design Contest

Jul 2020

- Awarded by: Chinese College Students Computer Design Contest Committee

Advanced Individual, 2018 - 2019 Around-Lake Marathon

Oct 2019

- Awarded by: Huazhong Agricultural University

SCHOLARSHIP

- Outstanding Study Scholarship (ratio: less than 0.1%) Jul 2021
- Excellent Study Scholarship (ratio: 10%) Nov 2021, Nov 2020
- Merit Student Scholarship (ratio: 20%) Nov 2020
- Advanced Study Scholarship (ratio: 3%) Jul 2020, Jul 2019, Apr 2019

TECHNICAL SKILLS

Programming Languages C, C++, Java, Python, R, Go, Scala, SQL, Verilog HDL

Operating Systems Microsoft Windows, CentOS, Ubuntu

Database Oracle, MySQL, SQLite, MongoDB

Frameworks Scikit-learn, Keras, TensorFlow, PyTorch, Hadoop, Spark

Tools Git, Docker