Yu Cheng

Room 668-13, Software Building, 2 Lushan S Rd, Yuelu District, Changsha, Hunan Province, China, 410082

📞 (+86) 182 0568 2382 | Email: Cjade@hnu.edu.cn

RESEARCH INTERESTS

My research interests lie in deep learning, especially deep generative models and reinforcement learning. Specifically, I am interested in applying deep generative models (VAEs, RNNs, GANs, Transformers & Flow-based models) in bioinformatics to design novel molecules for drug discovery. I am also interested in natural language processing and graph neural networks as two representation of molecules (SMILES, Gaph) is similar to human languages and graphs.

EDUCATION

Hunan University (HNU) | *Hunan, China*   Sept. 2020 – expected July 2023
Master of Computer Science and Technology, College of Computer Science and Electronic Engineering
- **GPA: 3.66/4.00**
- **Advisor: Prof. Zeng Xiangxiang and Prof. Song Bosheng (BIHAI Lab Information)**
- Coursework: Natural Language Processing Technology (A), Techniques and Applications of Knowledge Graph (A-) , Advanced Operating Systems (A-), Principles of Network Security (A+)

Anhui University of Science and Technology (AUST) | *Anhui, China*   Sept. 2016 – Jun 2020
Bachelor of Computer Science and Technology, School of Computer Science and Engineering
- **GPA: 4.02/5.00 (Top 0.88%)**

PUBLICATIONS (Google Scholar)

- **Yu. Cheng**, Jingxin Dong, Hongxin Xiang, Yuansheng, Liu, Bosheng Song and Xiangxiang Zeng, "DF-SGAN: Deep Fusion Generative Adversarial Networks for Image-Guided Molecular Sequence Generation", *In preparation, to be submitted to Bioinformatics 2022 (IF = 6.931/ JCR Q2).*
- **Yu Cheng**, Yongshun Gong, Yuansheng Liu, Bosheng Song, Quan Zou, Molecular design in drug discovery: a comprehensive review of deep generative models, *Briefings in Bioinformatics*, Volume 22, Issue 6, November 2021, bbab344, https://doi.org/10.1093/bib/bbab344. (Top SCI District 1, JCR Q1, IF=11.622, 2021)
- Bosheng Song, **Yu Cheng**, Uniform Solution to QAST Problem by Communication P Systems with Membrane Division and Promoters, Computer Science, 47.5 (2020): 38-42. (Recommended by CCF, B-level Chinese Periodical)

RESEARCH EXPERIENCE

De Novo Molecular Generation Research | *Hunan, China*   Sept. 2021 - Present
BIHAI Lab, Advised by Prof. Xiangxiang Zeng
- **Goal:** Generating novel molecules with certain desirable properties under two representations (Sequence and Image)
- Designed a generative adversarial framework consisting of a generator and discriminator. The generator is based on a sequence-based simplified molecular input line entry system (SMILES) and used as a stochastic policy for exploring chemical spaces. The discriminator is an image-based binary bioactive classifier which is used to encourage the generation of molecules that belong to the real bioactive-compound distribution
- Trained an LSTM-based generator via reinforcement learning and adversarial training to produce new molecules with high biological activities toward a given target. Beforehand, Pretrained such a generator on a large dataset ChEMBL to learn basic grammar of bioactive chemical space
- Proposed an image-based bioactive discriminator by first training the model (ResNet-18) using 8500,000 molecular images, and then added MLP and BN layers to train a bioactive classifier by finetuning it in an adversarial way
- Employed DF-SGAN to design molecules targeting dopamine receptor D2 and demonstrated that it can generate novel compounds with high similarity to the bioactive datasets in several structure-based metrics
- A paper is in preparation: DF-SGAN: Deep Fusion Generative Adversarial Networks for Image-Guided Molecular Sequence Generation

De Novo Molecular Generation Review | *Hunan, China*   July 2020 - Aug. 2021
BIHAI Lab, Advised by Prof. Yuansheng Liu, Bosheng Song
- **Goal:** Summarizing the recent advances of de novo molecular design for drug discovery and the current challenges in deep generative models for de novo molecular design

- Conducted literature reviews on the domain-related works and performed feature analysis on commonly used datasets such as ChEMBL, ZINC and QM9
- Divided existing state-of-the-art models into two categories according to molecular representations (Sequence and Graph) and then subdivided into four categories according to the four types of generative models (VAEs, GANs, RNNs, Flow-based models)
- Analyzed confronted challenges including data representation, bioactive data scarcity and black-box models
- Proposed potential solutions such as incorporating multi-omics data which makes up for the insufficiency of data scarcity
- A paper is accepted by Briefings in Bioinformatics

Sentimental Analysis on Weibo | *Hunan, China*   Mar. 2021 – Apr. 2021
Course Design
- **Goal:** Digging users' emotions to understand online news hotspots and monitor public fashion trend on Weibo
- Preprocessed the COAE2014 (around 5000) and loaded the pretrained RoBERTa-wwm-ext model (Pytorch)
- Finetuned the models on the Weibo datasets (Pytorch)
- Determined the emotional labels by testing on the data (Pytorch)
- Created a Flask & Bootstrap front-end interface to show the sentimental classification of Weibo text in real-time
- Received the Best Course Project award (Rank: 1/16 groups, based on the quality of work / technical achievement / Presentation)

RoboCupSoccer 3D simulation | *Anhui, China*   Sept. 2017 - Apr. 2019
Team Leader, National Second Prize of RoboCup China Open
- **Goal:** Pursuing the RoboCup 2050 Goal (team of multi-agent 3D humanoid robot soccer winning the champion human soccer team) by focusing on team strategy, humanoid robotics, optimization and artificial intelligence
- Led the research on how to employ the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) algorithm to successively generate and evaluate sets of candidate parameters regarding individual skills such as walking and turning (Matlab)
- Designed a series of related objective functions (also known as fitness measures) that returned a real number value indicating how well a candidate performed on the training task, once CMA-ES generated a group of candidates (Python, C++)
- Proposed a novel soccer strategy which is the first to introduce Hungarian algorithm in the assignment problem to solve the formation change of the whole team by a minimum cost during the competition (C++)
- Won the national second prize in RoboCup China Open-3D Simulation League 2019 (Top 3%)

SKILLS & OTHER QUALIFICATIONS

- **Programming**: Python (e.g., Pytorch/Scikit-learn/Numpy/TensorFlow), C++, Java Matlab, Latex
- **Media**: Editor of DrugAI (HIBAI Lab, Official Accounts, 30000+ Follows, Sharing the latest academic papers published on top conferences and journals like [Nature, Cell, Science])
- **Language**：Mandarin, English (IETLS, Band 6.5)
- **Art**: Folk Dance (Advanced, Ninth Grade/ Tenth Grade)

HONORS AND AWARDS

National Awards Scholarship, HNU & AUST	2021 & 2017
First-Class Scholarship for Excellent Students, HNU	2020
Academician Scholarship of AUST	2020
First-Class Scholarship for Excellent Students, AUST	2019 & 2018
RoboCup China Open, National Second Prize	2019
LanQiao Cup (Group C/C++), Received Provincial Second Prize	2019
Internet+ Innovation Contest (the Sixth Finding Cup), East China Division Third Award	2019