褚晏伊

在读博士生

可全职实习

1996 年 12 月，东北辽宁人

15941896746

a96123155@sjtu.edu.cn

上海交通大学

Education

• 2017-2022 年，上海交通大学，致远荣誉博士，生物信息学，GPA 3.83/4.0

• 2021 年出行，卡尔加里大学，联合培养博士，化学信息学

• 2013-2017 年，沈阳药科大学，学士，药物制剂，专业排名 TOP 5%

Skills

语言：Python, R, Matlab, SQL, Latex, Markdown, C++ 等

开发环境：Linux, Git, Shell 等

代码框架：DGL, Tensorflow, Keras, Pytorch, Networkx, Sk-multilearn, Sklearn, XGBoost, Requests 等

Interests

异质图网络 | 图神经网络 | 推荐算法 | 深度集成学习 | 多标签学习 | 生成模型 | 强化学习 | 数据挖掘

Experiences

本科：中国药学会科普志愿者、学院组织部长、学校下乡支教分队负责人

博士：学院研究生会副主席、学院社会实践团团长、VR 智能家居创业合伙人

Strengths

• 多领域知识储备
（AI 医药、推荐系统、NLP 等）
• 熟悉常用机器学习算法
• 对 AI 论文的快速理解
• 独立科研能力
• 善于合作

Project Experiences

2020.7-2020.9　光子工作室群　实习生　腾讯公司互动娱乐事业群，深圳

项目：应用图神经网络进行玩家流失预测

• 根据玩家个人信息、好友信息等建图，提取节点特征和边特征
• 构建多种图网络模型（GCN, GraphSAGE, R-GCN, GraphSAINT, GCMC, PinSAGE 等），同时对不平衡数据的处理方法进行优化（如过采样/降采样等）
• 模型性能达到理想结果（精度和召回率分别达到 0.9 和 0.6），效果可与使用复杂时序特征的 LSTM 相媲美

2019.1-2019.12 机器学习技术部　实习生　英特尔亚太研发有限公司，上海

项目：基于 MCTS 和人工神经网络的药物逆合成

我们首先利用深度生成模型扩大药物数据集，采用时下流行的图神经网络来自动化地提取药物特征。同时，为了解决数据集类别数量过大等问题，我们将标签平滑、AM-softmax 等技术首次应用于图神经网络中。接下来，本研究构建 AlphaGo 的类似模型返回药物逆合成的可行路径。

2018.10-2018.12　大数据中心 实习生　延安数莓信息科技有限公司，上海

• 豆瓣网站电影爬虫项目：独立完成，并做成可供同事使用的用户界面。
• 影视植入广告预测项目：独立完成，使用基于关联规则的推荐算法。
• 负责影视行业调研、评估影视项目的盈利可能，辅助上司的投资决策。

博士项目：应用人工智能进行药物发现和药物重定位

• 利用网络爬虫等技术构建新的数据集
• 利用图神经网络、自编码器、Word2Vec 等提取特征
• 模型：集成学习、深度学习、多标签学习、异常检测、社区检测、推荐算法等（详情见：https://github.com/a96123155/DTI-CDF 或 DTI-MLCD）

Publications

1. **Yanyi Chu**, et al. Predicting drug-target interactions using multi-label learning with community detection method (DTI-MLCD). bioRxiv preprint bioarXiv: 2020.05.11.087734. (2020)

2. **Yanyi Chu**, et al. DTI-CDF: A Cascade Deep Forest Model Towards the Prediction of Drug-Target Interactions Based on Hybrid Features *Brief. Bioinform.* doi:10.1093/bib/bbz152 (2019, **IF: 9.101**)

3. Mengyang Liu, **Yanyi Chu**, et al. Accelerated blood clearance of nanoemulsions modified with PEG-cholesterol and PEG-phospholipid derivatives in rats: The effect of PEG-lipid linkage and PEG molecular weight. Molecular Pharmaceutics. (2019, **IF: 4.396**)

Honors and Awards

• 2020 年，博士研究生国家奖学金
• 2020 年，先声药业奖学金
• 2020 年，第四届青年生命科学家论坛暨研究成果大赛三等奖
• 2018 年，全国大学生“互联网+”创新创业大赛校铜奖
• 2016 年，沈阳市优秀大学生
• 2015 年，全国大学生数学建模竞赛辽宁赛区三等奖
• 2015 年，上海信谊百路达药业有限公司企业奖学金
• 2013-2017 年，沈阳药科大学：优秀毕业生、十杰大学生、先进个人

Yanyi Chu

PhD Candidate

Full-time Intern

December 1996

15941896746

a96123155@sjtu.edu.cn

Shanghai, China

Skills

Languages： Python, R, Matlab, SQL, Latex, Markdown, C++, etc.

Development： Linux, Git, Shell, etc.

Frameworks： Scikit-learn, Tensorflow, Keras, Pytorch, Scikit-multilearn, XGBoost, Mlxtend, Networkx, Deep Graph Library, Requests, etc.

Interests

- Heterogeneous graph network
- Recommendation system
- Deep ensemble learning
- Graph neural network
- Multi-label learning
- Data mining

On-campus

- Vice-chairman of the college postgraduate students' union
- Head of the College Employment Social Practice Group

Strengths

- Multi-field Knowledge Reserve (AI medicine, Recommendation System, NLP, etc.)
- Good Understanding for AI papers
- Rich Experience in model coding
- Fast Learner
- Innovator

Education

2017-2022 Shanghai Jiao Tong University, China — Honorary PhD candidates
Major in Bioinformatics, GPA 3.83/4

2021(EXP) University of Calgary, Canada — Visiting PhD
Major in Cheminformatics

2013-2017 Shenyang Pharmaceutical University, China — Bachelor
Major in Pharmaceutics, GPA 3.5/4, TOP 5%

Project Experience

2020.7-2020.9 Lightspeed & Quantum Studios Group Intern Tencent

First, I build graph, extract node and edge features based on user's information. Then, construct a variety of graph neural network models (GCN, GraphSAGE, R-GCN, GraphSAINT, GCMC, PinSAGE, etc), and optimize the processing methods of unbalanced data (such as oversampling, downsampling, etc.). Finally, the model performance achieves the ideal result (precision and recall are 0.9 and 0.6, respectively), and the performance is comparable to that of LSTM using complex timing-sequence features.

2019.1-2019.12 Machine learning technology Intern Intel Asia-Pacific Co.

We work on *drug retrosynthesis which based on MCTS and deep learning.* First, the drug data set is expanded by deep generative model, drug features are extracted through graph neural network. Next, MCTS with deep learning are applied to return the feasible route for drug retrosynthesis.

2018.10-2018.12 Datacenter Intern Databerry

- Independently completed *the Douban website movie crawler project,* and made user software for colleagues.
- Use the recommendation algorithm based on association rules to independently complete *the project of film and television placement advertising prediction.*

PhD candidate project: Application of artificial intelligence in drug discovery and drug relocation

First, I use *data mining* techniques such as web crawlers to construct new data sets. Then use *graph neural network, autoencoder*, and *Word2Vec* to extract features. Finally, building classification models, such as *ensemble learning, deep learning, multi-label learning, outlier detection, recommended algorithms, etc*.

Honors and Awards

- 2020 National scholarship for Postgraduates
- 2020 Jiangsu Simcere Pharmaceutical Co., Ltd. Scholarship
- 2013-2017 Shenyang Pharmaceutical University：
  Outstanding graduates, Top ten outstanding college students, Advanced individual, Outstanding league member pacesetter
- 2016 Excellent college student in Shenyang
- 2015 3st Prize, National College Students Mathematical Modeling Competition Liaoning Division
- 2015 Shanghai Xinyi Bailuda Pharmaceutical Co., Ltd. Scholarship

Publications

1. **Yanyi Chu**, et al. Predicting drug-target interactions using multi-label learning with community detection method (DTI-MLCD). BioarXiv preprint BioarXiv: 2020.05.11.087734. (2020)
2. **Yanyi Chu**, et al. DTI-CDF: A Cascade Deep Forest Model Towards the Prediction of Drug-Target Interactions Based on Hybrid Features *Brief. Bioinform.* doi:10.1093/bib/bbz152 (2019, **IF: 9.101**)