

程中建

19171222812 ◇ 2838588360@qq.com ◇ <https://github.com/czjczj>

2022 应届硕士 ◇ [https://www.zjcheng.site\(blog\)](https://www.zjcheng.site(blog))

教育

2019.9 - 至今	中南大学	硕士 (保研)	计算机科学与技术	专业排名 top5 %
2015.9 - 2019.6	湖北大学	本科	软件工程	专业排名 top3 %

项目/ 研究经历

基于独立特征与交互特征的 DTI(药物 -靶标相互作用) 预测研究 2020.10 - 2021.2

项目介绍: 针对已提出的药物 -靶标相关作用预测模式的不同。构建一种模拟真实的药物 -靶标绑定过程, 具有生物学启发意义的交互特征提取结构, 该交互特征并融合药物和靶标的各自的独立特征预测药物 -靶标是否会相互作用。最终模型在多个数据集上分类效果超过 *SOTA*。

- 使用 GNN, 多核的 1D-CNN 结构分别提取药物和靶标的特征;
- 基于改进 Transformer 的 encoder-decoder 结构, 提取药物 -靶标序列表征中的交互特征;
- 数据预处理, 特征工程, 模型调参和优化器的选择。

成果: 一作论文, 已投稿 Bioinformatics (中科院 2 区)

基于多头自注意力和图注意力网络的 DTI(药物 -靶标相互作用) 预测研究 2019.11 - 2020.5

项目介绍: 针对当前基于 *DL* 的模型中对于药物和靶标的建模不足, 无法提取有效特征问题; 在研究中尝试使用 *GNN* 去表示和构建药物分子模型, 建立具有自注意力机制的 *Encoder* 结构。获取靶标序列中的上下文信息。最终通过特征融合进行分类, 在多个不同数据集上的分类效果超过 *SOTA*。

- 神经网络框架使用 Pytorch, 编程语言是 Python;
- 针对数据特点, 构建药物 -靶标特征提取器 (GNN, Encoder based on multi-head self-attention)
- 尝试过其他序列相关的神经网络结构 (包括 LSTM, RNN, CNN) 等。

成果: 一作论文投稿在 IEEE/ACM Transactions on Computational Biology and Bioinformatics(CCF B 类期刊), 已一审返修。

农业问答数据处理挑战赛 2020.7 - 2020.9

比赛介绍: 农业大数据中存在许多的专业名词, 如农作物、病虫害等信息。比赛提供互联网平台上的专家与农民的问答数据, 要求构建模型, 对问答数据进行标记切词, 初赛排名 (11/586). 复赛排名 (20/586)

github: <https://github.com/czjczj/AgriNER>

- 数据清洗, 模型构建, 调参。
- 尝试了 HMM, CRF, Bi-LSTM, Bi-LSTM+CRF, 最终选用 Bert+Bi-LSTM+CRF 效果更好。

知乎问题自动标注 2020.6 - 2020.7

比赛介绍: 给定一个知乎问题, 标注出该问题所属于的多个标签。每个知乎问题对应的标签个数不相同。最终提交模型得分第一。

github: https://github.com/czjczj/Zhihu_MultiLabel

- 改进 bert-base-chinese 的下游任务, fine-tune 预训练模型。
- 尝试了 Seq2Seq+Softmax 等多分类的模型, 效果没有 bert 好。

技能

- 编程语言：熟悉 Java, Python; 了解 C++/C 语言。
- 机器学习：熟悉 numpy, matplotlib, sklearn 和 xgboost; 熟悉 Pytorch 神经网络框架; 了解 tensorflow。
- 技能证书：软件设计师 (中级), CET-6(具有良好的英文读写能力)。
- 其他方面：熟悉数据结构、计算机网络。了解 Oracle, MySQL 和 Linux 基本操作命令。
- 学术论文：
 - 1) 2021.2: IIFDTI: predicting drug target interaction through interactive and independent features based on attention mechanism[J]. Bioinformatics. (一作, Under Review 已投稿, 中科院二区)
 - 2) 2021.3: Zhao Q, Yang M, **Cheng Z**, et al. Biomedical data and deep learning computational models for predicting compound-protein relations[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics. (二作, Accept, CCF B 期刊)
 - 3) 2020.5: Drug-target interaction prediction using multi-head self-attention and graph attention network[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics. (一作, Major Revisions 已返修, CCF B 期刊)

荣誉/ 奖项

- 国家级：
 - 国家奖学金
 - 国家励志奖学金
 - 蓝桥杯软件设计大赛国家二等奖
 - 全国大学生数学建模竞赛国家二等奖
 - 全国大学生信息安全竞赛决赛三等奖
 - 中国软件杯全国大学生软件设计大赛决赛三等奖
 - 美国大学生数学建模竞赛 H 奖
- 省部级：
 - 蓝桥杯软件设计大赛湖北省一等奖
 - 全国大学生数学竞赛湖北省三等奖
 - 华中地区数学建模竞赛三等奖
- 院校级：
 - 校级科研创新奖学金
 - 校级一等学业奖学金 (3 次)
 - 校级三好学生 (2 次)

自我评价

- 性格开朗乐观, 善于与人交际;
- 有责任心和上进心, 勤奋刻苦, 对待工作认真负责, 不拖延;
- 有求知欲, 具备良好的自主学习能力和问题解决能力;
- 具备良好的团队合作精神, 在新环境中具有较强的适应能力;