

李洁厅

http://www.hellojet.cn

求职意向：自然语言处理算法工程师

Email : lijieting@zju.edu.cn

Mobile : +86-18668420278

毕业时间：2021.03

EDUCATION

- 浙江大学 杭州, 浙江
计算机学院 2018.09 – Present
- 电子科技大学 成都, 四川
软件工程学院 2014.09 – 2018.06

EXPERIENCES

- 医学知识开放平台 2019.09 - Present
 - 背景: 针对医疗文献中存在大量的临床路径知识不能高效利用和传播, 以NLP、检索为技术基础的医学知识开放平台对这些知识进行抽取整理, 形成计算机可解释文档, 方便医生临床使用和知识共享。
 - 负责:
 1. 采用预训练模型+BiLSTM+CRF和HanLP词典匹配相结合的方法进行医疗实体抽取, 完成系统的自动标注模块; 通过半监督和主动学习的方法对药品说明书进行实体抽取, 针对数据量少造成的边界不清问题进行不同数据源的混合加权训练;
 2. 针对医疗实体链接数据的缺少, 回标产生数据, 使用别名表和elasticsearch模糊匹配产生候选实体, 基于BERT构建pointwise ranking链接模型;
 3. 对CIG和医疗信息标准化和交互等进行调研, 分析医疗本体库, 筛选临床有关, 基于cnSchema设计并验证医疗知识图谱schema。
- 基于鲸鱼群优化的社区发现算法 2017
 - 背景: 社区是复杂网络中连接紧密的子网络, 挖掘这样的社区, 即社区发现, 对社交网络兴趣圈挖掘、中药相互作用探索等有积极意义。启发式算法对于社区发现问题有着不错的效果, 但是算法速度不佳。本工作将标签传播融入初始化, 改进启发式算子, 保证实验效果和算法速度的提升。
 - 负责:
 1. 阅读社区发现相关文献, 对现有应用于社区发现的元启发式算法进行分析总结; 复现AFSA、BA、IMeme-Net、DPSO算法代码;
 2. 修改WOA算法(Whale Optimization Algorithm)的启发式算子, 设计新的标签传播初始化策略, 适应社区发现问题, 在合成网络上最优, 真实网络上大部分最优。

CONTESTS

- 面向中文短文本的实体识别与链指 CCKS 2019
排名: 8/349 2019.04 – 2019.07
 - 背景: 任务是识别中文短文本中的实体, 并与给定知识库中的对应实体进行关联。难点在于短文本口语化严重, 上下文信息较少。本题分为实体识别和实体链接两个过程。
 - 负责:
 1. 实体识别: 使用字词混合向量+双向LSTM、DGCNN模型, 构造分词、词典、位置等特征, 与队友的预训练模型进行模型融合; 对训练集进行蒸馏去噪;
 2. 实体链接: 对识别和链接进行联合建模和两阶段建模, 用双向LSTM+maxpool层抽取语义特征, 构造上下文特征、类型特征和热度特征等。
 3. 撰写评测论文: https://conference.bj.bcebos.com/ccks2019/eval/webpage/pdfs/eval_paper_2_3.pdf

● 金融信息负面及主体判定

CCF 2019

初赛: 11/1447, 复赛: 17/1447

2019.08 – 2019.11

- 背景: 从海量的金融文本中快速准确地挖掘出关键信息对投资者和决策者具有重要意义。本赛题关注金融文本中信息主体的负面消息检测, 分成句子的情感判断和实体的情感判断两个过程。
- 负责:
 1. 调研aspect-level情感分析论文, 实现基于aspect2context attention的情感分析模型, 设计实现主体级和句子级情感分析的多任务学习;
 2. 针对数据过长进行截断, 清洗电话网址等脏数据, 设计数据输入bert的形式; 在训练集和测试集上继续进行BERT预训练, 增加模型多样性, 制定融合策略和后处理策略

● 基于Adversarial Attack的问题等价性判别

DIAC 2019

排名: 29/1146

2019.11 – 2019.12

- 背景: 针对专业性较强的问答系统(如法律、政务等), 准确的判别用户的输入是否为给定问题的语义等价问法是智能问答系统的关键。本次任务在问题等价性判定的基础上引入错别字同音字重复词等对抗样本进行对抗攻击, 要求用户的模型具有鲁棒性。
- 负责:
 1. 采用InferSent模型、acrossAttention模型、Bert模型和triplet loss等进行句子相似性判断;
 2. 使用同义词替换、随机交换、删除tf-idf值较低的词、翻译, 随机mask等方法进行数据增强, 实现FGM对抗训练。

SKILLS

- **Languages:** Python, C++
- **Frameworks:** Keras、tensorflow
- **Knowledge:** 机器学习 (LR、SVM、决策树、EM、最大熵、HMM), 自然语言处理 (分词、情感分析、句子匹配)、知识图谱 (命名实体识别、关系抽取、实体链接、知识图谱嵌入)
- **English:** CET-6 504

HONOR

- **研一:** 浙江大学优秀研究生
- **本科:** 多次人民奖学金和企业奖学金, 排名11/73