

奉雨娟

手机: (+86)18801016306 邮箱: 2426781817@qq.com

性别: 女 出生日期: 1993 年 5 月

求职意向: AI/Healthcare、自然语言处理、数据挖掘



教育经历

清华大学	计算机科学与技术	博士(直博)	2015/09-2020/07
	核心课程:	统计机器学习, 统计推断, 高级算法等	人工智能所
天津大学	计算机科学与技术	学士学位	2011/09-2015/06
	核心课程:	人工智能基础, 算法设计分析, 概率论与数理统计等	GPA: 3.79 (1%-5%)
康奈尔大学	康奈尔威尔医学院	访问学者	2019/04-2019/10
	核心课程:	医学信息, 数据挖掘, 大数据分析, 机器学习	

项目经历

多模态医疗健康大数据信息融合和表示学习	2019/04-2019/10
内容: 基于多模态时序数据 (ECG 信号、生命体征、病历文本等) 预测 ICU 病人死亡率风险	
贡献: 数据挖掘会议论文第一作者; 实现数据集清洗, 预处理, 特征提取, 模型预测及评价框架; 实现时序数据特征提取 (如 ECG: CNN-LSTM; word2vec 词向量训练等); 提出基于 Memory Network 的数据融合算法, 解决多模态、异步时序数据采样频率差别大、异质性强等问题。	
结合知识图谱的医学概念的 embedding 学习及其应用	2017/01-2017/07
内容: 学习病案首页中医学概念 (ICD10 和 ICD9 编码) 的 embedding 向量, 结合 ICD 编码系统结构图谱, 学习病人的表示向量, 预测病人住院总费用。	
贡献: 会议论文第一作者; 提出 MG-CNN 模型, 将 ICD 编码层级结构信息和 CNN 网络结合; 用 Glove 预训练医学概念的词向量及 T-SNE 可视化; 提高低频率医学概念表示质量。	
ICU 病人的急性肾损伤 (AKI) 风险预测任务中的类别不均衡问题研究	2018/03-2018/07
内容: 基于 ICU 病人的历史病历数据, 提前预测未来发生 AKI 的风险, 实现早期诊断和疾病预防; 解决由于类别不均衡 (患病人数远小于非患病人数) 影响模型准确率的问题。	
贡献: 会议论文共同第一作者; 实现已有的不均衡采样方法 (over-sampling, under-sampling); 提出基于病人相似度采样构造均衡数据集的方法; AKI 鉴别的影响因子分析;	
科学文献摘要检索任务-Precision Medicine Track 比赛	2019/05-2019/08
内容: 输入患者的特定信息 (疾病名称, 基因变异和人口统计学信息), 返回 Pubmed 文献数据库中最相关且属于精准医疗范畴 (涉及疾病的治疗、预防、预后等方面) 文章摘要。	
贡献: 参赛小组成员; 基于整理好的疾病和基因的知识图谱, 拓展疾病、基因变异的同义词集合; 用 Bert 预训练得到检索词、所有摘要的向量表示, 根据相似度排序检索结果; 用 BOW、tf-idf、Bert 词向量等特征, 训练机器学习模型, 判断文章属性; 小组选为口头报告展示。	

实习经历

基于 VGG 预训练的图像检索任务	清华大学天津电子信息研究院	2017/06-2017/08
内容: 鉴于日常图像、医学影像中图像检索需求, 输入给定图片, 返回数据库中最相似图片		
贡献: 个人独立完成; 利用 ImageNet 上预先训练 VGG 提取图片的表示向量; 利用近似最近邻算法 (ANN) 返回最相似图片集合, 相比最近邻算法 (Nearest Neighbor) 提高了检索速度;		
健康医疗病案首页数据分析	江苏省卫生统计信息中心	2016/07-2016/08
内容: 分析江苏省数十家医院的病案首页数据, 利用统计机器学习方法探索 “病人流向分布分析”, “血型 and 疾病的关联分析”, “疾病诊断与环境、天气、空气质量的关系” 等问题。		
贡献: 小组成员; 承担 “血型 and 疾病的关联分析” 任务, 负责数据处理、模型实现;		

课程助教

- 高级算法设计与分析 陈挺 （清华大学计算机系教授，国家千人计划） 2017-2018 春季学期
统计学方法及其应用 江瑞 （清华大学自动化系副教授） 2016-2017 秋季学期

论文发表

1. Yujuan Feng, Zhenxing Xu, Lin Gan, Ning Chen, Bin Yu, Ting Chen, Fei Wang, **DCMN: Double Core Memory Network for Patient Outcome Prediction with Multimodal Data**, IEEE International Conference on Data Mining (ICDM), 2019. (CCF B, 9.08%).
2. Yujuan Feng, Xu Min, Ning Chen, Xiaolei Xie, Haibo Wang, and Ting Chen, **Patient Outcome Prediction via Convolutional Neural Networks based on Multi-Granularity Medical Concept Embedding**. In Bioinformatics and Biomedicine (BIBM), 2017. (CCF B, 19.0%).
3. Zhenxing Xu*, Yujuan Feng*, Yun Li, Anand Srivastava, Prakash Adekkanattu, Jessica S. Ancker, Guoqian Jiang, Richard C. Kiefer, Kathleen Lee, Jennifer A. Pacheco, Luke V. Rasmussen, Jyotishman Pathak, Yuan Luo, Fei Wang, **Predictive Modeling of the Risk of Acute Kidney Injury in Critical Care: A Systematic Investigation of The Class Imbalance Problem**. AMIA Summits on Translational Science Proceedings, 2019. (共一, CCF C).
4. Zhenxing Xu, Yujuan Feng, Lin Gan, Xu Min, Anais Chaland, Ning Chen, Bin Yu, Ting Chen, Fei Wang, **Double Core Memory Network for Mortality Prediction using Multimodal Intensive Care Data**, Oral Presentation for KDD Workshop on Applied Data Science for Healthcare, 2019.

科研项目

1. 医学人工智能产品检测共性关键技术及标准研究（课题 4），国家重点研发计划项目，2019/11-2022/10，520 万元，参与，在研
2. 单细胞组学数据异质性解读的计算理论和方法，国家自然科学基金面上项目，2019/01-2022/12，66 万元，参与，在研
3. 中文电子病历自然语言处理和智能诊断，清华-福州数据技术研究院项目，2018/07-2020/06，80 万元，参与，在研
4. 面向医疗健康大数据分析的高效隐层空间学习方法研究，国家自然科学基金面上项目，2017/01-2020/12 63，参与，在研

学校实践

- 清华大学生活委员（2015-2017）：管理班级开支，积极组织男女生节和班级建设。
山西平遥暑期社会实践（2012/07）：作为实践小组组织者，策划和开展山西省平遥古城文化的传承的调研实践活动，获得暑期实践报告一等奖和暑期社会实践先进个人。
志愿服务实践（2011-2015）：参与第九届全国大学生运动会 100 天启动仪式、善行壹佰爱心包裹活动（金额 700 元），被评为中国扶贫基金会一星级志愿者，志愿服务先进个人。

获奖情况

- 清华大学博士生短期出国访学基金（2018/11），国家奖学金（2014/11，2012/11），金龙鱼奖学金（2013/11），北洋励志奖学金（2011/05）
天津大学优秀毕业生（2015/01），天津大学年度三好学生（2014/11，2013/11），自强不息先进个人，学习优异先进个人，（2013/11）

个人技能

- 编程语言：python 熟练，c++、linux 一般；深度学习框架：keras 熟悉，tensorflow 一般；
外语水平：CET-4:591，CET-6: 518，CET-4/6 口语:B
自我评价：性格沉稳，做事态度认真，责任心强；有较强团队合作精神和沟通协调能力。