

李琪斐

手机：(+86) 18857448068 · 邮箱：wxliqifei@sina.com



教育背景

北京航空航天大学, 北航复杂数据分析研究中心, 工业工程硕士 2018.09 - 2021.06

- GPA: 3.7/4.0, 相关课程: 信息检索, 数据挖掘, 计算机网络, 程序设计, 计量金融等

法国里昂中央理工学院 (法国 Top5 工程师学校), 双学位通用工程硕士 2017.09 - 2019.06

- GPA: 3.8/4.0, 相关课程: 数据分析, 机器学习, 算法与数据结构, 测度论与随机过程, 数值分析, 概率论, 统计等

北京航空航天大学, 数学与应用数学学士 2014.09 - 2018.06

- 专业前 5%, 相关课程: 线性代数, 数学分析, 概率统计, 计算机图形学, 电磁学, 波动物理, 数学物理, 量子物理等

实习经历

微软亚洲研究院 2021.01-2021.04

研究实习生, 社会计算组 北京, 中国

- 基于深度学习和自然语言处理的新闻推荐, 参与了多个项目的模型训练调优、结果分析和模型上线, 将实验所用的 tensorflow 代码重构成用于上线的 pytorch 版本, 并参与论文的修改。

摩根大通集团 2020.07-2020.12

机器学习实习生, 量化部门 北京, 中国

- 基于神经网络和其它深度学习模型对金融工具的 PnL 进行预测, 探索模型在衍生品定价方向的约束和局限性, 并给出用于衍生品定价的神经网络特征定义和训练样本选择的通用指南, 模型取得了优秀的线上表现, 得到项目组的高度评价。

中国农业科学院 2019.10 - 2020.05

数据分析实习生, 质标所研发部门 北京, 中国

- 创造性采用全新指标体系, 将多次采样下的大规模农药残留稀疏矩阵进行降维, 建立统计模型分析; 运用熵值-Topsis 法、灰色关联度法、变异系数法等综合评价方法构建指数, 实现对年度全国范围内各类蔬菜农药残留风险评估。
- 该项目构建的评价指标功能在绝大多数蔬菜品种上可以代替传统指标的功能, 同时可发现一些传统方法难以识别的风险, 在农药残留安全领域上运用统计学方法首次实现识别蔬菜农药混合污染风险。

BIVOLIS 2018.05 - 2018.09

软件开发实习生, 研发部门 巴黎, 法国

- 根据公司需求设计并建立一套新的系统, 用于进行电影数字包的检验以及实现发送过程自动化
- 使用 python 和 Linux 进行软件开发, 设计的系统已在企业投入使用, 并为企业带来很大效益

科研及项目经历

Learning from Perturbations: Diverse and Informative Dialogue Generation with Inverse Adversarial Training

- 该论文已被 ACL 2021 接收 (共同一作)
- 该论文提出了一种用于训练对话系统的新的“逆向”对抗训练方法。该方法通过鼓励模型在接受被扰动的对话历史输入的情况下不能输出同样的对话回复, 提高了对话系统的对话历史建模能力并且可以有效减少对话系统生成无聊回复的现象。

Connecting the Dots Between Fact Verification and Fake News Detection

- 该论文已被 COLING 2020 接收 (Oral, 第一作者)
- 该论文提出了一种新的 zero-shot 的假新闻检测方法。该方法利用预训练的文本摘要模型将待检测的新闻文章压缩成较短的摘要, 基于 AllenNLP 提取摘要关键词输入 Google 搜索引擎提取证据, 之后利用预训练好的事实检测模型在提取的外部证据的支持下对该摘要的可信度进行判断, 从而判断原

始新闻的可信程度。该方法成功地将事实检测领域的进展迁移到假新闻识别任务中，使得无监督的 zero-shot 假新闻检测成为可能。

Learning to Detect Persona for Dialogue Personalization without Explicit Persona Description

- 该论文投稿于 EMNLP 2021（第二作者）
- 该论文提出了一种新的对话系统个性化的方法。该方法通过训练一个以对话历史为输入，预测对话参与者人格特征的人格检测模型，从而在推断阶段实现不依赖对话者人格信息的对话系统个性化。

DebiasRec: Bias-aware User Modeling and Click Prediction for Personalized News Recommendation

- 该论文投稿于 NeurIPS 2021（第四作者）

CCF 大数据与计算智能大赛—技术需求与技术成果项目之间关联度计算

- 将“项目需求”与“项目成果”文本作为句子对，考虑不同节选输入的方法将截断后的句子对输入 BERT 模型，采用交叉验证对模型进行训练，最后输出技术需求与技术成果关联度。
- 最终 score 为 0.7833/1(baseline 为 0.7522)，大赛排名 62/862。

学生日常运动轨迹数据可视化分析

- 以里昂中央理工学生为元素建立一个具有代表性的采样样本，用 Googlemap 对样本进行 GPS 位置信息的收集，对影响 GPS 精度的绝大多数因素进行分析；基于 k-means 等算法，用 python 对数据实现可视化并分析预测。
- 该项目实现：基于学生及教职工经常停留的地点对学校教学楼和功能楼的重新排置提出建议和意见，基于用户过去已知位置信息预测下一时刻的位置点，分析影响 GPS 精度的绝大多数因素，对学校附近的公交站给出合适的时刻表。

社团和组织经历

里昂中央理工学校国际办公室	2017.12 - 2019.02
Bureau International 副主席	里昂, 法国
<ul style="list-style-type: none">• 由 10 个来自不同国家的学生组成的学生组织，旨在帮助国际学生融入校园生活和促进国际文化交流• 作为活动负责人策划并主办赴德国慕尼黑啤酒节庆祝活动，中国周，世界美食交流活动，各种晚会等	
里昂中央理工学校大型春节联欢晚会	2018.02
宣传总负责人	里昂, 法国
<ul style="list-style-type: none">• 在 Facebook 等社交网站以及校园内进行活动宣传，春晚宣传影像策划负责人	

技能/证书及其他

技能:	Python(熟练), R(熟练), matlab, Latex, Adobe Premier Pro, Adobe lightroom
语言:	法语 (DALF C1 证书 (法国教育部法语能力考试, 接近母语水平)), 英语 (CET6, TOEIC 865/990)
证书:	机器学习 (Coursera 证书), 神经网络和深度学习 (Coursera 证书)
爱好:	滑雪, 旅行, 健身, 摄影, 视频剪辑
奖项:	北航校优秀研究生 (2019), 北航学业奖学金一等奖 (2019, 2020), 北航校优秀毕业生 (2018) 北航校优秀团员 (2016), 北航校优秀生 (2015), 北航校学习优秀奖学金一等奖 (2015)