

刘翔

15617865777 | 846458320@qq.com
1999-02 | 汉族 | 团员



教育经历

香港大学

计算机科学 硕士 工程学院

2022年09月

香港

乔治梅森大学

计算机科学 本科 工程学院

2018年08月 - 2022年06月

美国维吉尼亚州

- GPA : 3.75/4.00
- 主修课程：自然语言处理，计算机图形视觉处理，数据结构，数据库管理
- 多次进入院长名单

河南大学

网络工程 本科 软件学院

2017年09月 - 2022年06月

开封

- GPA : 88/100, 年纪排名4/292
- 参加中美国际121人才交流计划，在美国梅森大学学习交流三年，2022年6月本科毕业，获得双方本科学位。

工作经历

百度研究院认知计算实验室

实习生

2021年12月 - 至今

北京

- 调研OIE，逻辑推理，语言层次化方向论文，复现论文代码。
- 完成OIA(A Predicate-Function-Argument Annotation of Natural Language for Open-Domain Information eXpression)项目在中文语料上的部署，根据中文语法特点制定相应的标注规则、处理规则、转换规则等。在中文数据集上OIA也可以发挥接近英文数据集的效果。
- 独立完成了基于WikiCoref数据的共指问题，模型需要在同一簇的指代关系中选出最具有代表性的提及。使用了基于图的拓扑排序、基于置信度的累加排序、LightGBM的list wise排序。
- 个人收获：可以更加高效的阅读论文，学习论文方法并思考如何使用论文方法解决当前项目。更快速的上手大型项目，大幅度锻炼了工程能力。增强了与同事和老师的沟通能力，有问题的想法都要及时沟通。

项目经历

微博疫情谣言预测模型 | 独立负责人

2022年03月 - 2022年04月

- 通过使用清华NLP实验室提供的疫情相关谣言数据和中文谣言数据，使用Pytorch建立了基于BERT, LSTM, AutoGluon的疫情谣言预测模型。完成了数据探索，数据的处理，词向量特征处理，BERT, LSTM和AutoGluon模型的搭建，模型训练，指标的展示，以及反思和改进。
- 在这次项目中，个人经历了从提出想法，寻找数据，参考论文，代码实践，模型搭建和文章的撰写的全过程，对NLP项目的流程有了非常深刻的理解。

CCF个贷违约比赛TOP5% | 组长

2021年09月 - 2021年11月

- 赛题任务: 利用用户个人信息和消费信息预测是否违约。
- 赛题难点: 使用内部两个有差异的数据进行共同建模。
- 使用了lightgbm和xgboost建立单模模型，最终使用模型融合stacking和权重融合技术来提高成绩。
- 学习了特征工程在工业项目中应用，更加了解特征工程对于机器学习的重要性；掌握了lightgbm和xgboost两种boosting树模型的使用方法，以及optuna调参技巧。

情感分析案例 | 独立负责人

2021年07月 - 2021年09月

- 在影评文本数据集上进行情感分析，手动实现了词袋模型和N-gram模型的特征提取方式。
- 并且根据hinge loss损失函数推导出对应的梯度下降函数，手动实现梯度下降。
- 对不同的模型进行调参，分析模型泛化能力以及参数对模型性能的影响。
- 通过这个项目，更加深刻的理解了深度学习的原理，了解了梯度下降的真正用法，并且对于基于统计的语言模型有了深刻的理解。

个人总结

计算机双学位，对于中外的计算机知识都有很好地掌握，既拥有中式学习要求的知识强度，也拥有外国院校对知识的强烈求知欲和团队工作模式。

主要语言：Python, 有使用经验：C++, C, Java, JDBC, SQL

热爱算法，对机器学习和深度学习有较大兴趣

出色英文水平，可以阅读英文文献，了解最新的研究方法