

我的姓名

联系电话：136***** 现居住地：深圳 期望地：深圳 期望职位：自然语言处理工程师

性别：男 出生日期:1994.12 QQ: 83***** 电子邮箱：*****@outlook.com

教育背景

墨尔本大学 | 工程学院 | 信息技术 硕士

2018.02-2019.12

马来西亚博特拉大学 | 经济与管理学院 | 经济学学士

2012.09-2016.07

专业技能

- 熟悉常见的机器学习算法和自然语言处理模型，如 RNN, LSTM, 有 BERT 应用经验
- 编程语言：Python, Java, Haskell, Prolog
- 深度学习使用平台：Pytorch
- 有特征工程经验如特征挖掘，特征选择等
- 能无障碍阅读英文文献和技术文档

项目经历

搭建事实验证系统 (自然语言处理相关)

2019.04-2019.06

- 项目简介：该项目基于维基百科语料库，构建自动验证语句系统。当输入一条待判定语句，系统会根据语料库自动验证该语句是否为真、假或无法核实，并提供语料库中的具体句子证明有效性。
- 具体做法：(1) 文档提取：旨在提取出与待验证语句相关的文档。先对语料库文档和待验证语句进行预处理，用 PyLucene 库对文档建立索引，以便高效率地搜索与待验证的语句有高相关性的文档；(2) 证据抽取：从提取出的相关文档中抽取出来能用来判断待验证语句有效性的句子证据。分别尝试了 BERT 的两种模型，即分类模型和问答模型。分类模型主要是判断文档中的每条句子与待验证语句是否相关，如果相关的话就提取出来；问答模型则把待验证语句当成一个问题，并在文档中提取相应答案；(3) 事实验证：当提取出来一句或多句相关句子（答案）后，再训练一个 BERT 分类模型，判断待验证语句应为真、假，如在文档中提取不出相关句子（答案）则判断待验证语句为无法核实。
- 项目结果：该系统在 Codalab 竞赛中排名前 10%，F1-score 71.2%（为使用 BERT 分类模型的结果，问答模型结果稍逊）。
- 总结：(1) 主要难点一是训练前对数据的预处理，才能准确从维基百科语料库中搜索出与待验证语句相关性高的文档；(2) 难点二是 BERT 提供的问答模型对每个训练样本仅支持一篇文档，一个问题，零或一个答案，而该项目的训练样本是多篇文档，一个问题，答案可能会有多个。针对这个难点，我特意发邮件咨询当时在斯坦福 SQuAD 问答榜单上获得第一名的哈工大 NLP 组，在其建议下，我修改部分代码，使 BERT 问答模型可以对多个答案进行建模训练，最后根据预测出的答案的概率分布，抽取一个或多个答案；(3) 该项目仅仅使用了单个 BERT 模型，还有提升空间，如使用 XGBoost，随机森林等 ensemble 类方法提升精度。

吃豆人竞赛 (强化学习，路径规划相关)

2019.08-2019.11

- 项目简介：伯克利大学经典的强化学习和路径规划项目，设计算法使吃豆人吃到比敌方更多豆子从而赢得比赛。
- 主要任务：应用了 A-star 搜索，蒙特卡洛树搜索，值函数估计等算法，规划吃豆人的进攻和防守套路。
- 项目结果：竞赛第二名。
- 总结：该项目主要锻炼的是对版本控制 (GitLab) 的应用和对算法的理解，每天都有比赛进行，过程中需要针对算法不断优化迭代。

搭建基于推特事件监测系统 (论文复现，算法相关)

2018.11-2019.02

- 项目简介：基于推特发帖的定位，搭建一个从时间和空间监测事件发生地点的实时系统。
- 主要任务：在时间上用泊松分布模型检测事件，在空间上用四叉树定位事件地点。
- 项目结果：用 python 的 Tweepy 库接入推特的 API，实时获取本地推特帖子，成功检测出了当时在附近发生的春节活动和烟花秀。

自我评价

- 自学型（勤能补拙型）：不希望过于依赖别人，遇到问题想办法解决，一遍不会再来一遍，动手实操。常用的知识获取来源有：(1) 谷歌：主要搜索 Stack Overflow，技术文档，Medium，GitHub 等；(2) 知乎：搜索相关问题，都会有很多答主在分享（不乏很多大佬）；(3) 百度：主要搜索国内技术博客等；(4) 斯坦福人工智能资源站：如 CS224N 自然语言处理和 CS224U 自然语言理解等；(5) Youtube，B 站：海量学习资源，各种知识讲解，教程，公开课等。(6) 书籍：如《统计学习方法》。
- 团结型：不是独行侠，喜欢跟志同道合的小伙伴们交流，互助，分享，合作，这样才能提高效率 and 生产力。如我曾在一个全员对 APP 开发零基础的团队中担组长，通过快速自学，开发出 app 的基础版，再协调全员对 app 进行迭代更新。
- 对工作的态度：高效完成基础工作，然后在完成的基础上寻求完美，不断地迭代优化。