

个人简介

本人拥有扎实的数学和计算机基础，熟悉机器学习、深度学习相关概念，模型和算法，对自然语言处理有很大的兴趣。在三年以上的实践经验中，能熟练处理和分析数据，解决复杂的业务问题，具有快速学习和高效处理问题的能力。

教育经历

荷兰埃因霍芬理工大学 (Eindhoven University of Technology)

理学硕士学位

2016-09 - 2018-03

- 专业: 数据科学工程 (提前 6 个月以优异成绩 (cum laude) 毕业)
- GPA: 8.42/10.0 (前 5%)
- 毕业设计: Concept Drift Investigation(概念漂移调查) 分数:9/10

北京航空航天大学

工程学士学位, 理学学士学位

2012-09 - 2016-06

- 专业: 计算机科学与技术, 应用数学二学位
- GPA: 3.22/4.0 (前 30%, 优秀毕业生)
- 毕业设计: 基于树分解的交通异常子网检测 (优秀论文)
- 荣誉奖项: 美国大学生数学建模竞赛二等奖、北航博彦奖学金二等奖、北航冯如杯科技作品竞赛三等奖

工作经历

杭州观远数据有限公司

算法工程师, 项目团队负责人

杭州

2018-05 - 至今

- 主要参与联合利华销量预测项目, 从 POC、正式合作以及项目交付的整个流程, 2019 年开始成为项目负责人, 管理、协调团队资源
- 根据业务对数据进行清洗, 使用机器学习模型 (例如 LGBM, RNN, Prophet) 对数据进行建模、预测未来三个月的销量
- 负责大陆 3 个品类 (总共 4 个) 的预测建模, 负责联合利华台湾的整个项目
- 成功达到了项目一期的预测准确度要求 (比人工高 10%), 目前进行项目的落地
- 主要创新点: 使用自定义损失函数和特殊的多模型融合策略弥补树模型预测值偏保守的缺陷

荷兰合作银行 (Rabobank)

数据科学实习生

乌得勒支 荷兰

2017-07 - 2018-05

- 根据官方文档学习 H2O, 实现了一篇论文的方法, 并且完成了一些欺诈检测的实验
 - 使用 pauc 评价模型, 保证 FPR 一定的情况下得到较高的 TPR
 - 使用基于聚类的方法, 不同的聚类结果类别使用不同的模型
- 2017 年 9 月开始进行毕业设计, 针对欺诈检测和钓鱼邮件检测两个业务场景提出了一种新的概念漂移检测算法 (DDEAL)
 - 两个业务场景都不能实时获取数据的标签, 算法中提出用数据里的一个特征来做代理标签
 - 多模型融合进行概念漂移的检测, 将该算法与现有的一些已有的算法通过实验进行对比
 - 为业务专家开发了一个可视化应用来与概念漂移检测结果进行交互, 进一步定位变化位置

新境影像(创业项目)

算法合伙人

埃因霍温 荷兰

2017-07 - 2018-02

- 设计并开发算法对摄影人博客、摄影文章 (英文) 进行全网范围的搜索和爬取。成功获取到数千个摄影人博客以及其中数万篇文章
- 开发爬虫爬取国内摄影交流网站 (例如: 图虫网、摄影之友、微信公众号等)。
- 对爬取到的中文文章搭建摄影相关词库, 并进行 word2vec 建模, 实现摄影领域的词的联想功能
- 利用阿里云的推荐系统服务搭建文章、摄影人的推荐系统

项目经历

Kaggle Competitions Expert(目前排名前 2%)

2019-03 - 至今

- 桑坦德银行用户交易预测 top 6%
 - 模型: LGBM 和 CNN
 - 使用数据增强丰富训练集, 使用 Focal loss 解决正负样本比例失衡
- 性别代词解析 top 12%
 - 模型: Bert 加全连接层

iPhone 8 谣言分析

2016-09 - 2016-11

- 与另外四名成员共同参与该项目, 负责推特爬虫以及谣言真假的预测
- 开发一个基于 Python Scrapy 和推特 API 的爬虫, 用来根据关键字爬取相关推文, 手动给数据打标签
- 数据清洗, 预处理分词, 使用 word2vec 对单词进行建模, 多模型融合 (SVM, KNN, NB, LR) 进行文本分类
- 训练集基于 iPhone 7 的数据, 迁移应用到 iPhone 8 的测试集上, 成功预测了下一代 iPhone 上将会取消 home 键

个人技能

编程语言 Python, R, Java, C#, HTML, C++, C, PHP, JavaScript, SQL

职业技能 多任务处理、快速学习、数据分析、熟练阅读和理解英文论文和技术文档

语言技能 英语精通, 荷兰语入门