

江之源 (Alex Jiang)

1991.06.29 江苏无锡人 现居新加坡

(+65) 86791519

alex44jzy@gmail.com

alexjiangzy.com/cv

正直, 专注, 积极, 实干, 团队精神, 海外求学工作经历。2 年开发工程师经验, 熟悉前后台的敏捷开发流程, 善于分析业务和技术的结合点解决实际问题。对数据科学和分析建模有很强的敏锐度, 对机器学习和深度学习有一定的积累, 对自然语言处理 (NLP) 相关领域有很大的研究热情。双硕士, 国际视野, 现攻读新加坡国立大学 (NUS) 企业商务分析硕士学位, 新加坡 IBM 实习数据科学家。

求职意向: 数据科学家, 机器学习算法工程师, NLP 工程师, 意向地点: 上海。

教育背景

2018.01 - 2019.07 新加坡国立大学 (新加坡) 企业商务分析机器学习, 硕士

- GPA: 3.95
- 主修课程: 商务分析基础, 数据分析, 决策与优化, 文本挖掘, 新媒体与情感分析, 机器学习与文本处理, 用户关系管理 (CRM), 概率图挖掘等。

2013.09 - 2016.06 西南大学 (211, 重庆) 信号与信息处理, 硕士

- GPA: 4.0, 导师段书凯教授, 研究方向: 忆阻器件, 神经形态系统等
- 荣誉: 连续两年研究生科研奖学金, 校三好研究生, 西南大学研究生优秀毕业生。

2009.09 - 2013.07 中北大学 (太原) 电子信息工程, 学士

工作经历

2018.08 - 2019.04, IBM, 新加坡, 数据科学家 (实习)

- 项目 IBM Q2C 供应链时间序列预测及聊天机器人优化, 基于 IBM Bluemix Watson 系统, 实现供应链监控垂直领域 FAQ 的问答管理。
- 研究冷启动情形下文本相似度加权得分排名 (句嵌入, 词向量平均, 莱温斯坦距离, 余弦相似度等), 设计阈值用于直出答案或用户相似问题推荐。
- 研究多轮对话中意图分割语境重置问题, 埋点记录用户手动点击重置作为 Label, 选取 3 轮次对话构建基于 Tensorflow 的 WordSeq GRU 模型, 验证集 F1-score 73%。
- 单日全球用户订单数时间序列预测, 进行数据收集、清洗、划分及整合为时间序列。模型优化采用 Season ARIMA+LightGBM, 目标值与 Seasonal ARIMA 预测的残差作为 LightGBM 预测目标。特征方面优化, 手动构建历史特征及多组特征, 爬虫抓取订单主要国家节假日 (强特), 预测验证集结果 MAPE 较基线降低 13%。
- 基于 Flask, Gunicorn 实现模型与代码部署, 提供 RESTful API 用于线上 Chatbot 应用请求。

2016.07 - 2018.01, Coupang, 上海, 软件研发工程师

- 参与 Coupang 全网长途车辆管理分析平台产品线 4 个阶段 (OA-可视化-预测-模拟) 前后端整体迭代开发。全网长途车辆定位及可视化, 前后端使用 Angularjs, Google Map API 和 Java 微服务实现物流长途车辆可视化、数据分析平台及实时消息推送等。
- 结合配送路程时间, 准点率, 制定监控 KPI 分析长途路线健康状态, 并与 Vendor 端集成 Slask 实现消息提醒。针对不同路线车辆装载率不稳定容错能力低的问题, 开发模拟系统生成所有路线的最优派车方案, 与实际业务协调发掘实际派车问题, 有效降低长途物流链路成本。
- Hub 仓储手持设备应用开发, 基于 React Native, Redux 和 Realm 组件化开发安卓 PDA 手持扫码设备应用程序, 实现对包裹信息的扫码录入、错误提醒和日志跟踪。基于服务端解析和 CodePush 热更新, 实现版本发布及灵活热修复模式, 前往韩国仁川仓储中心参与实地部署。

2015.06 - 2015.09, 英语流利说, 上海, 算法工程师 (实习)

- 实现用户音频数据与朗读文本序列化, 基于 Scala 参与开发英语流利说 Spark 大数据自动化打分系统 Pipeline, 并使用 Apache Zeppelin 配置、输出各版本打分结果及各音素细节可视化, 多语种含噪声基准模型自动化验证等, [Demo](#) 收录在 ZEPL 展示案例中。

科研及比赛

2018.10 - 2018.12, 阿里巴巴天池瑞金医院 MMC 人工智能辅助构建知识图谱大赛 TOP 2%, 第二赛季 37 / 1629, 第一赛季 78 / 1629。

- 第一赛季实体标注, 基线模型 BiLSTM+CRF, 通过调整隐含层层数超参训练 3 个不同模型, 对预测的 3 组标注结果取并集 Voting 融合, 融合结果 F1-score 较单模型提升 2%线上最高 72%。
- 第二赛季关系抽取, 基于阿里 PAI 平台, 2 大类 10 组关系 (疾病类 5 组关系, 药物类 5 组关系), 采用实体对齐截断语句作为训练样例的远程监督, 为防止训练中正负样本比例失调, 正样本跨窗口长度截取 1500, 负样本构建以 2 大类实体标注为中心结合统计结果, 前后窗口分别截断长度 800 和 1100。模型基于 Pytorch 采用 LSTM+TextCNN 实现, 后将截断语句进行头尾补齐后, 复现 PCNN (Piecewise) 进行分段池化有一定提升, 线上 F1-score 63.4%。

2018.09 - 2018.11, 基于新加坡 MyCareersFuture 挖掘分析新加坡 AI 人才需求特点及现状

- 使用 BS4, Selenium 抓取全网含有关键字的非结构化招聘信息。使用分词、正则规则、POS tagger、Ngram 词频解析分类提取职位关键信息, 基于 TF-IDF 和 LDA 挖掘生成不同维度的类别主题 (工作内容, 职业要求等)。

2014.01 - 2016.05, 参与国家自然科学基金 (61372179) 忆阻器件/CMOS 混合结构的智能 PID 控制器研究

- 第 5 届国际信息科学与技术国际会议 (IEEE ICIST) [发表论文](#) A threshold adaptive memristor model analysis with application in image storage. Information Science and Technology (ICIST), 2015 5th International Conference on. IEEE, 2015 (EI 检索)。

2018.03, 新加坡建屋发展局 (Singapore HDB) 二手组屋多维度销量、价格、区域化趋势分析的 [Tableau 可视化 Dashboard](#)。

2015.01, 清华大学信息科学与技术国家重点实验室 (无锡) 第 7 届国际物联网和云计算大会, 合作导师刘云浩 (IEEE Fellow), 参与国家自然科学基金面上项目 (61672436) 基于多模感知和移动互联网协作的导盲系统关键问题研究。

2014.07, 第 9 届华为杯研究生电子设计大赛二等奖, 智能门窗管理系统, 通过 ARM7、手机 GSM 模块及外部传感器实现自动化门窗开关。

专业技能

- 编程语言: Python, Java, Javascript, 了解 Bash, 函数式编程 Scala。
- 机器学习: LR, DT, NN, SVM, 降维, 聚类, 贝叶斯网络, 概率图模型, 模型融合提升, 时间序列分析, 优化方法等。
- 深度学习 (NLP): Tensorflow, Keras, Pytorch, Gensim, StanfordNLP 等; 熟悉词嵌入、Doc2Vec 原理, 了解 BERT 原理; CNN, RNN 模型 (GRU, LSTM 等) 原理; 关注文本分类, NER, 文本相似度分析, 知识图谱构建, Chatbot 等。
- 数据分析工具: Numpy, Scikit-learn, Pandas, Matplotlib, R, Tableau, Google Analytics Tool, 图挖掘 Gephi 等。
- Web 工程开发: Python (Flask), Java (MVC), Javascript (ES6, Angularjs, React Native), 数据库 (MySQL, MongoDB) 等。
- 其他: 英语听说读写熟练, 企业商务分析 (CRISP-DM), Agile, Git 管理等