

赵鹏

联系电话: 18676732070
邮箱: zp913913@gmail.com
期望城市: 深圳

Github: <https://github.com/tade0726/>
知乎: <https://www.zhihu.com/people/Pon9/>

个人评价

从事数据挖掘相关工作3年, 在客群分析、工业仿真、数据处理、数据挖掘、数据采集、自然语言处理有丰富的实践经验

目前想寻找数据挖掘工程师/机器学习算法工程师的岗位工作

工作经历

深圳大疆创新科技有限公司, 数据分析师 - (2018.04-2018.08)

- 维护开发 Jupyter 数据分析平台/为可视化报表系统开发接口/改造平台 Python 项目的日志收集
- 维护公共数据维度库, 爬取清洗整理全球国家经济/社会/地理数据
- 对销售渠道数据进行挖掘, 挖掘城市的销售潜力

深圳华策辉弘科技有限公司, 数据挖掘工程师 - (2015-2018)

- 服务于顺丰航空物流事业部, 开发基于 Python 的物流仿真系统, 负责开发技术栈的选型, 负责核心代码编写, 最终仿真的匹配度达到95%以上
- 服务于平安证券, 开发舆情分析平台, 负责编写自然语言处理模型, 用于预测 app 短评的情感倾向以及提取评论的主题关键词
- 开发爬虫系统, 爬取新浪微博/大众点评/房天下
- 服务于香港赛马会, 进行客户分群, 客户行为挖掘分析
- 作为数据挖掘技术 lead, 为公司引进 Python 数据挖掘生态, 对内部进行培训宣讲, 对外部客户进行宣讲培训

教育背景

- 深圳大学, 理学学士, 应用数学专业 - (2010-2014)
- Coursera, 机器学习课程证书 - (2015-2016)
- Udacity, 机器学习工程师纳米学位 - (2018-2019)

项目经验

深圳大疆创新科技有限公司, 渠道门店数据分析与 PYTHON 开发

- 爬取线下零售商店地理位置信息, 清洗整理并进行规范化
- 电商数据分析挖掘: 挖掘出电商数据存在转运海外的订单, 挖掘出存在高校集中购买无人机的行为
- Jupyter 客户端组件功能开发, 添加 ES SQL Query 功能, 对客户端进行重构优化
- 维护 Django 项目, 对 api 进行改造, 支持请求数据的翻页功能
- 开发基于 scrapy 爬虫项目, 爬取线下商圈的地理信息

顺丰航空物流事业部, 物流中心仿真项目

- 作为技术领导, 引进了完整的开发测试流程, 工具链, 制定工程师招聘标准, 编写技术文档, 制定开发计划时间表
- 对原有的仿真代码进行重构, 引进 SimPy, 对于仿真代码的业务逻辑进行优化, 降低了开发工作量, 缩短了仿真运行时间, 提高了仿真系统的通用性, 运行时间由原本26小时优化到20分钟
- 添加行业商业仿真软件所缺失的功能, 添加了机器开关状态/人力资源进行的动态调整, 引进了排班表的概念, 模拟了生产环境的真实情况

项目经验

平安证券舆情 NLP 模块开发

- 开发 NLP 模块, 实现 app 评论的情感倾向预测, 评论的主题提取
- 情感分类模型经过三次迭代后, F1-score 由 88 提高到 93
- 设计了反映证券 app 业务的文本主题体系, 利用 word2vec 挖掘相似主题, 替代传统的人工检阅或者机械正则匹配
- 对 app 评论进行新词挖掘, 利用开源工具 AutoParse, 自动提取有价值的词组
- 实现词向量的可视化挖掘, 利用了 T-sne 对词向量进行降维, 实现词向量在 2D 的可视化
- 尝试使用 LSTM 神经网络预测情感, 由于数据量较小, 效果比 word2ec 方案较弱, 没有进入生产环境

香港赛马会BI咨询

- 对不同来源的数据, 进行清洗关联, 构建基于客户 id 的维度数据表
- 进行特征工程, 根据客户业务特点, 构建上百个衍生变量
- 建立聚类模型 (K-means), 进行客户分群, 挖掘不同群体的行为特征, 识别客户群体忠诚度, 对高价值用户/潜在可提升的客户/流失客户进行标记
- 生成具有解释性的群体划分指标, 利用决策树对分群结果进行规则的提炼, 生成的规则指标覆盖聚类 90% 以上的客户, 用于描述 K-mean 聚类的结果
- 编写了一个类, 自动化组合生成候选的衍生变量, 提高工作效率

技能

- 数据分析: Python/Pandas/SQL/Spark/Hive/Jupyter
- 工业仿真: SimPy
- 自然语言处理(NLP): gensim/jieba/snownlp
- 机器学习/DL: Scikit-learn/Kera/Tensorflow
- 机器学习算法: K-means/SVM/Decision-Tree/Logistic/word2vec/T-sne/PCA
- 网络爬虫: Scrapy/lxml/bs4/pyspider/
- Web框架: Django/Tornado
- 数据库相关: MySQL/PgSQL/SQLite/MongoDB/Elasticsearch/Redis
- 容器、版本管理、文档和自动化部署工具: Svn/Git/Markdown/Docker

项目和作品

项目

- 物流中心仿真项目: 顺丰中转场仿真项目对物流分拣中心的物件分拣效率进行模拟, 基于 Python SimPy
- 零售商场地理信息爬取项目: 爬取万达/顺电/苹果等线下零售商铺的地理信息, 基于 Python Scrapy + Mongoddb

技术文章

- [Python SimPy 仿真系列文章 \(1\)](#)
- [Python SimPy 仿真系列文章 \(2\)](#)