

“Stay hungry, Stay foolish.”
- Steve Jobs

专业技能

编程语言 **Python = Shell > Java > C++**, 扎实的算法基础.

机器学习 **熟悉word2vec、textcnn、rnn、Transformer以及Bert等常用NLP技术**, 熟悉CF、LR、SVM等常用机器学习算法, 熟悉tensorflow框架.

Hadoop **熟练掌握MapReduce编程**, 了解Hadoop运行机制及架构, 熟悉Hive, 了解HBase.

版本管理 **熟练使用git和svn等版本管理工具**.

英语 **能用英语交流**, 熟练阅读英文技术资料.

工作经历

2016.03 - now **杭州绿湾技术有限公司**, 策略工程师.
大数据系统开发, NLP相关项目开发

2014.03 - 2016.03 **百度**, 策略工程师.
新闻推荐系统的开发

2010.07 - 2014.03 **风行视频技术有限公司**, 数据挖掘工程师.
数据仓库、推荐系统的开发

项目经历

2019@绿湾 **NLP - 短文本分类**.
使用word2vec得到词向量, 利用300篇人工标注数据, 使用Bi-GRU+CRF实现提取人名、机构名、时间、POI等属性的识别.

2019 **NLP - 属性识别**.
使用word2vec得到词向量, 利用300篇人工标注数据, 使用Bi-GRU+CRF实现提取人名、机构名、时间、POI等属性的识别.

2018 **NLP - 中文信息抽取**.
通过查找资料, 学习实践开发通用的中文信息抽取工具。主要命名实体识别、实体关系抽取和事件信息抽取三个功能.

2018 **NLP - 人物常驻地挖掘**.
从库中导出事件信息, 抽取关注的身份证、时间、地点等字段, 对人物、时间、地点、事件优化级、事件频率做统计, 使用自举方法完成常驻地的挖掘.

2018 **NLP - 法院文书信息抽取**.
通过使用规则抽取法院判决书中的人、身份证号、被告/原告、时间等信息, 用于知识图谱建模及后续挖掘.

2016-2018 **大数据分析系统 - 分发层开发**, 知识图谱, Java.
使用Spring Boot框架, 完成分发层代码开发.

2015@百度 **推荐系统 - 质量模型训练**, 内容模型.
人工标识一堆文本, 评价分为高中低三档, 分别训练高质量分类器(高/中低)和低质量分类器(高中/低), 然后对每篇文章打上高质量分和低质量分, 用于推荐排序及过滤.

- 2015 **推荐系统 - 新词发现, 内容模型.**
通过人工标注色情新闻级别得到语料, 使用词袋模型和LR模型建立一个分类模型, 根据新闻标题和内容做色情度评定, 用于解决CTR预估中色情新闻过多的问题
- 2014 **推荐系统 - 更新新闻分类器, 内容模型.**
通过人工标注新闻分类得到最近新闻分类语料, 与之前的语料合并作为训练、测试与验证集, 使用TFIDF计算权重, 使用libsvm训练新闻分类器, 解决因分类器太旧造成的分类错误。最终将分类准确率提升至93%
- 2014 **推荐系统 - 新闻色情度分类, 内容模型、评分打压过滤.**
通过人工标注色情新闻级别得到语料, 使用词袋模型和LR模型建立一个分类模型, 根据新闻标题和内容做色情度评定, 用于解决CTR预估中色情新闻过多的问题
- 2013@风行 **推荐系统 - 视频聚合影视推荐, 关联推荐、冷启动.**
通过抓取外网媒体信息 (包括媒体简介、主演、导演、标签等) 计算媒体间的相似度, 给用户做关联推荐, 用于解决冷启动问题。
- 2013 **推荐系统 - LR模型对推荐结果二次排序, LR模型、二次排序.**
采用用户反馈数据 (隐式 (观看、点击、浏览) + 显式 (顶、踩、打分等)) 进行LR建模, 对CF的推荐结果进行二次排序, 优化效果
- 2013 **数据仓库 - BIEE数据仓库迁移, MapReduce、Kettle、BIEE.**
将数据仓库从原来的单机使用脚本及Infobright计算, 迁移至Hadoop使用MapReduce, 以提高系统的可扩展性和稳定性。同时建模和前端使用BIEE, 数据库采用Oracle。数据连接及导入使用Kettle, 通过数据分析与管理平台进行任务调度和管理。
- 2012 **数据仓库 - 数据分析与管理平台, Hadoop、Hive、PHP、MySQL、Python.**
主要使用PHP+Python+MySQL实现一个统一管理风行所有上报数据、ETL数据以及调度任务的系统。目标是通过该系统, 方便业务数据有管理、任务调度, 并能开放整个Hadoop计算资源和数据资源给其他部门, 降低各部门使用数据的门槛。

管理经历

- 2012 - 2013 **Scrum敏捷开发**, 数据BI团队的Scrum master, 负责召开各种需求评审会、迭代会、回顾会等, 带领团队完成BI组各类项目。
- 2012 - 2013 **研发主管**, 负责BI组的人员招聘、绩效考核、团队建设、人员培养等工作。

教育经历

- 2007.09 - 2010.07 **硕士**, 北京工业大学, 计算机软件与理论。
- 2003.10 - 2007.07 **本科**, 南昌大学, 计算机科学与技术。