

关注

机器学习 ◆ 深度学习 ◆ 人工智能 ◆ 目标检测 ◆ 后端系统 ◆ 逆向工程 ◆ 数值优化

技能

Python	Keras, Pytorch	MongoDB, Redis
Javascript	jQuery, Vuejs	Numpy, Scipy, Cython
HTML5	Flask, Sanic, Curio	Scikit-learn, OpenCV
CSS3	Bootstrap, Element-ui	Amazon AWS, Aliyun

学历

数学与科学计算(本科)

复旦大学

上海, 中国

2002 - 2006

工作经历

索来科技(自雇)

技术总监, 产品助理, 架构师, 后端, 前端, 算法研究, 算法实现, 数据清理, 数据库设计, 系统运维, 配置装机, 网络管理

2015.9 - 2019.2

公司由我和另外两个朋友联合创立, 先后搞过很多小规模但难度比较大的项目.

比如沪牌人工智能竞拍, 金融量化交易平台, 股票和邮币卡打板, 银行一体化羊毛系统等.

期间, 作为技术主管, 主导了全部项目从顶层设计到具体实现的每一个步骤. (小公司招的人不太给力T.T)

- 沪牌验证码识别 - 深度学习 - 目标检测 - SSD / Yolo
- 沪牌反侦测系统 - 可以理解为反-反外挂
- 沪牌可扩展高并发调度与处理架构
- 沪牌策略回测、拍卖系统模拟
- 沪深股市Level2全数据下载 - 逆向工程 - 异步处理
- 沪深股票回测系统 - 比JoinQuant和RiceQuant更为准确(无论是数据还是算法实现上), 还实现了基于Level2数据逐笔回测
- 沪深股票/邮币卡抢打板系统 - 抢连续涨停股的第一笔买入 - 并发300个阿里云ECS - 逆向工程 - 云系统管理
- 羊毛综合系统 - 整合银行U盾, 手机接码, 第三方系统逆向 - 参加羊毛活动, 然后高效变现汇集

菜根科技

技术总监

2014.3 - 2015.9

菜根科技是一家小创业公司, 规模为30-6人.

CEO曾是德意志银行的VP, 理念是挖掘商业数据中的隐藏信息来指导企业决策.

- 研究使用TDA(Topology Data Analysis), 一个比较冷门的传统Machine Learning的某个黑科技
- 研究使用关联性分析(Association rule learning)来解决商品推荐等问题
- 开发与情报监控系统, 爬虫 - ETL - 分析
- 各种数据可视化, 数值计算优化

Favbuy

技术总监

2012.8 - 2014.3

Favbuy是一家小创业公司, 规模为20人, CEO曾是SEM达人.

公司一开始打算做一个时尚潮牌特卖的聚合网站, 通过SEM来引流抽成.

失败后转方向做数据挖掘, 巅峰时监控全淘宝天猫2亿件商品的销量和评价信息.

- 电商大数据: 500台ecs做分布式爬虫 - 20台ecs做数据库 - redis做数据索引和清洗 - 自定义的map-reduce架构
- 时尚特卖: 爬虫 - 数据清洗 - 机器学习分类

虎扑体育

游戏后端技术负责人

2011.11 - 2012.8

虎扑体育是国内垂直体育类网站, 当时规模为300人左右

我加入的游戏团队, 为了开发一款名为《虎扑世界》的网页游戏

期间还顺便帮助虎扑的后端团队解决了高并发体育文字直播的架构问题

- 虎扑世界: gevent + mongodb + flash架构的页游 - 用观察者模式解决战斗难题
- nba直播: 文字直播翻译 + nodejs高并发(3w人同时在线)

cchars(未完成开源)

一个通用的图片字符识别系统

2018-2019

- CNN + GRU + CTC-loss, 按行扫描的识别模型
- SSD: 魔改VGG和Anchor的Detection模型
- Yolo: 加入汉字和自定义图片的魔改版Yolo

sinal2 <https://github.com/observerss/sinal2>

新浪Level2数据客户端

2017

- 下载Level2的新浪数据
- 异步处理+多核

ngender <https://github.com/observerss/ngender>

根据中文姓名来猜测这个人的性别, 当时项目需要, 所以顺手做了一个开源出来

2015

- 基于Naive Bayes的人名猜测
- 核心代码只有10行

textfilter <https://github.com/observerss/textfilter>

超快的关键字过滤

2012

- 基于确定有限状态机(DFA), 过滤速度和过滤样本库无关
- 和GFW用的是同一个原理

simplecd <https://code.google.com/p/simplecd/>

一个简易的电驴资源爬取的网站, 时值强监管, verycd资源大量被删, 于是做了个备份, 在当时算是小有名气的项目

2009

- python+sqlite的极简架构
- 爬虫全无框架, 硬撸