

姚忠辉



男 | 1994 年 10 月生 | 户口：山西运城 | 现居住于深圳南山
Tel:18819046110 (微信同号)
Email: 1325546813@qq.com
Github: <https://github.com/huber-yaoer>
知乎用户名：姚忠辉

求职意向

- 期望职业： spark 机器学习工程师
- 工作地区： 深圳、北京、上海
- 目前状况： 办理离职中

教育经历

2012/09 --2016/06 太原工业学院 | 电子信息工程 | 本科

专业技能

编程技能 |python\C++\java\scala、spark/hadoop 生态圈、Numpy\Scipy\Pandas\scikit-learn、qt\pyqtgraph、linux、TensorFlow\Keras\PyTorch、mysql\MongoDB\Redis\Neo4j

数学基础|高数、线代、概率论、统计学、优化、信息论、傅里叶变换\拉普拉斯变换\FFT、矢量分析与场论、微分方程
机器学习理论|Lass\ridge\感知机\LR\LDA\SVM\BP、决策树\随机森林、adaboost\GBDT、朴素贝叶斯、PCA、聚类、关联、CNN、RNN、LSTM、遗传算法、强化学习、GAN

知识来源

部分书籍：

《图解数据结构》 《西瓜书》 《统计学习方法》 《数据挖掘导论》 《深度学习（花书）》
《hadoop 权威指南》 《spark 与 hadoop 大数据分析》 《数字图像处理-冈萨雷斯》 《国富论》
《Python 数据分析与数据化运营》 《企业大数据系统构建实战：技术、架构、实施与应用》

部分网站与视频：

万门大学：高等数学特训班、线性代数特训班、概率统计特训班、傅里叶变换、拉普拉斯变换与小波变换、微分方程特训班、矢量分析与场论特训班、 人工智能、大数据与复杂系统一月特训班

吴恩达：《机器学习》、《深度学习》

邹博-小象学院-机器学习升级版 II

邢不行 - Python 量化入门

西安电子科技大学-任荻荣：数字图像处理 MIT 线性代数课程精细笔记

微信公众号号搜索：马同学高等数学、遇见数学、SPSS 生活统计学

王福重经济学-腾讯 V 课堂

Kaggle 数据集

知乎搜索姚忠辉了解更多，特别关注这两篇文章：

自己整理机器学习思维导图

大数据学习笔记

2018/08 -- 至今 康美健康小镇投资有限公司 岗位：机器学习工程师

公司简介：

康美健康小镇投资有限公司，2017 年 6 月份注册成立，是康美实业集团公司旗下新成立的全资子公司。

康美实业是大健康领军企业康美药业股份有限公司的控股股东，主要从事实业投资、资本投资、房地产开发等业务。

康美健康小镇投资有限公司，投资项目响应国家医改红利政策，整合“大健康+大平台+大数据+大服务”体系的中医药全产业链精准服务型“互联网+”资源，将大健康产业作为特色小镇的产业核心，聚集健康养生养老、文化体验、休闲度假、体育运动、健康管理、健康产品、旅游娱乐、农业种植等产业业态，打造完善的产业体系，旨在于全国布局多个各具特色的“候鸟式智慧养生健康文化旅游项目”。

所在大数据组（属深圳总部运营部）工作内容：

大数据平台 CDH 搭建与维护，收集小镇各业务系统数据、物联网平台数据、网上爬取数据，挖掘分析，精准营销。

基于图数据库 Neo4j 构建“康美知识图谱”，内容包括资产管理、健康评估与建议、营养膳食、运动建议、药品信息、旅游信息等。

基于知识图谱构建问答系统供“健康管家”使用；

基于知识图谱和数据中台（包含有个人基本信息、健康信息、交易信息），构建个性化推荐系统；

参与产品设计及数字化运营；

本人工作内容：

团队内主要负责数据分析、算法选型。

知识图谱及推荐系统找腾讯云、阿里云、达摩院协助解决，我是这部分工作技术接口人，负责 demo 编写，推进项目、把控方向。

大数据平台搭建及运维、ETL、数据爬取；

参与产品设计及运营；

辅助精算师对小镇会员量预测进行模型搭建（用聚类、凸优化进行计算机辅助运算）；

腾讯觅影技术接口人；

2016/04 -- 2018/04 深圳市中钞科信金融科技有限公司 岗位：软件工程师

公司简介：

深圳市中钞科信有限公司成立于 2008 年 8 月，由中国印钞造币总公司、中国科学院、深圳信达投资公司三方合资组建，专业从事高端银行机具设备、机器视觉检测设备、系统集成与数据服务，是国家级高新技术企业和软件企业。公司总部与研发基地设在深圳、机器视觉检测设备的研发和生产基地设在成都、清分机产品生产制造基地设在沈阳。

所在软件组介绍：

主要负责（钞票行业）行业外项目，核心竞争力是外观质量检测，涉及领域包括：制卡厂产品线各阶段卡的外观检测、OCR 及距离测量，还包括手机玻璃、电池极片、优衣库标签、光纤等的外观检测，概括一下就是做机器视觉。

本人工作内容：

软件研发：负责交互层（python+qt\pyqtgraph+twisted）架构设计及实现。

其它：驻厂分析客户真实需求，提出算法改进意见，交付产品，跟踪项目进展，给新客户演示设备。

获优秀员工奖和项目奖。

项目经验

1. 电商平台推荐系统及数字化运营（康美健康小镇）

项目目标：对会员数据、商品数据、订单数据及埋点数据进行收集过滤存储，经过 ETL 操作后进行离线分析统计，构建事实标签，模型标签，预测标签，进而实现**个性化推荐及数字化运营**。

整个项目架构包括：使用 cdh 搭建高可用的集群，flume 进行数据的采集与预处理，生产数据到 kafka 集群，kafka 集群对数据分类存储，hbase 消费 kafka 中数据，进行海量存储，hive 使用 hbase 的数据作为源数据，再对 hive 中数据用 sparkSQL（python）进行数据清洗，根据需求做相应的 hive 表设计，而后：

对于**数字化运营：**sparkSQL 调用 hive 进行 ETL 操作实现 RFM 等模型，sqoop 导出 RFM 等模型输出数据到 mysql，前端调用 mysql 中的数据进行展示分析，azkaban 作为整个系统任务调度的工具。

对于**推荐系统：**利用 sparkML 运行 ALS 模型进行训练，训练完成，利用 ALS 为会员进行商品推荐，商品推荐保存到 hive，sqoop 导出推荐信息到 mysql，前端调用 mysql 中数据进行商品推荐。

注：ALS（交替最小二乘法）常用于基于矩阵分解的协同过滤推荐系统中。

责任描述：**主要负责数据分析、机器学习部分，也参与平台搭建，ETL 等**

2. 卡外观质量检测系统（中钞科信）

简介：设备分软件、电气、机械三部分，软件分为交互层、算法层，交互层用 python，算法层用 C++，交互层是中心节点，可通过网络调用算法层及控制硬件，实现了对卡片油污、脏点、漏印、划伤、色差、图文偏移等外观问题及 OCR 的实时检测，并将好坏卡分类装盒。系统一秒可处理 7 张卡片。系统可支持十个以上工位，每个工位功能不同，要求识别缺陷多种多样，系统可识别 95% 的缺陷。

后期总结了卡外观检测系统的主要问题和视觉检测常用需求，开发了**新框架**，相对旧版框架，**功能更多，代码更少，主体框架一目了然，高内聚、低耦合、复用性更强**。在新框架上扩展的业务有：电池极片、光纤、信封、芯片，优衣库标签等的外观检测。

获奖情况：被深圳软件行业协会评定为**深圳市 2017 年优秀软件产品**

业绩：在制卡厂领域，国内处于垄断地位，竞争对手为德国纽豹集团，国外客户包括奥地利、美国、越南

主要编程工具及手段：**python, qt\pyqtgraph, protobuf, twisted, numpy, opencv, 密度聚类等**

收获：**代码规范，深刻理解 python 之禅。（因为也是集群，这阶段对 hadoop、spark 做过调研）**

责任描述：软件部分**两个人**，我主要负责交互层，参与框架设计，经历了整个产品研发周期，多次按照客户要求迭代，也负责测试、驻厂、反馈算法检测效果并提出改进意见。

3. 量化交易初试

简介：股票择时策略：**双均线**；读数据，计算复权价，计算均线，找出买入卖出信号，计算仓位，考虑跌停涨停、滑点、手续费、印花税，上市一年后开始买卖，计算资金曲线。

股票选股策略：**小市值**；选股数据整理，每月市值排名，选取市值最小的十只股票，计算资金曲线，策略评价指标

股票获取数据及实时交易：**tushare** 或直接爬取新浪财经，数据整理，自动登录中信建投账户买卖股票，远程邮件通知

基于 **LSTM** 的比特币**多因子**交易策略

主要编程工具：**python, pandas, matplotlib, keras**

自我评价

代码规范，深刻理解 python 之禅，python 是母语，而后依次是 C -> C++ -> java -> scala

电子信息工程出身，编程语言从 C/C++ 开始学起，更能理解整个计算机科学内在逻辑；

毕业后重新学了一遍数学；

熟悉常见机器学习算法；

了解 hadoop、spark 生态系统各个组件的适用范围及特性；

乐观豁达，执行力强，数学爱好者，终身学习者；

热爱爬山、单车穿越、滑板等运动；