

# 林祥熟

年龄：28

联系电话：15068729148

[linxiangshu@126.com](mailto:linxiangshu@126.com)

<https://maples.me>

<https://github.com/y521263>

## 个人技能

熟悉 C++/Python/Golang 编程，有较好的代码风格和编程习惯；

熟悉风控业务算法，预测、图像识别，爬虫抓取等相关技术；

熟悉常用机器学习算法，掌握回归、分类等算法模型并解决实际问题；

熟悉深度学习原理，熟悉 TensorFlow 深度学习框架；

熟悉Hive/Hadoop等大数据开发工具；

对技术有热情，能自我驱动，通过学习解决困难问题；

## 工作经历

### 云集 高级算法工程师

2019.04-至今

风控：基于文本分类的异常地址检测模型(Fasttext)，识别黄牛收货地址，效果显著；

风控：构建全平台的用户风险评分体系，结合用户下单，设备，注册，活跃数据，产出风险系数；

风控：基于spark Graph计算构建图关系网络，以用户id，设备，支付账号，收货手机为主体，采用社区发现算法(LabelPropagation)挖掘出黄牛社群；

预测：物流仓库每日单量预测，大量梳理活动数据，结合仓库发货逻辑做特征工程，效果：仓库达标率从30%提升至50%；优化人力安排，降低成本；

预测：平台销售额预测，结合当日活动情况，采用多类目预测增加样本数量，预测准确率93%，为运营决策提供支持；

预测：仓库Top 10 sku预测，输出仓库每日销量Top 10 的sku；效果：60%的仓库，覆盖率40%；

### 盈盈理财 资深算法工程师

2018.05-2019.04

风控：参与消费分期风控模型的开发与优化，主要采用xgboost算法，结合业务数据建模；深入基础数据源，不断优化特征工程；

- 针对各种数据源，建立不同风控模型，提升模型精准度

- 不断深入数据源，优化特征工程，提升AUC值2个百分点

**预测：**平台每日理财交易金额预测；根据当日回款、运营活动等数据，再基于SVM的回归模型，建立预估模型；

- 梳理多样的运营活动数据和业务逻辑，不断优化特征工程；
- 交易金额预估精度控制在200万以内，极大提升运营效率；

### **talkingData-AdTracking-Fraud-Detection (Kaggle top 10%)**

Kaggle上关于talkingData 点击反欺诈检测比赛；构建丰富特征的同时，需要解决亿级的数据计算问题；小样本调参，全量数据微调，加快迭代速度。

- 受计算资源限制，积极回收内存，每完成一个feature再dump到磁盘
- 本地小样本cv，加快迭代速度，优化模型效果
- 不同时间窗口，连续点击间隔等统计值，优化特征工程

### **挖财 算法工程师**

2014.06-2018.05

**图像识别：**验证码识别，基于目标检测的理论基础，采用 Faster-RCNN 算法，实现 End-to-End 识别逻辑，适用于各种类型图片；单张图片检测耗时控制在150ms左右，准确率达到0.95。训练集为4万张左右

- 开发与维护样本数据集(4万+)
- 不断调优，优化选框策略，持续提升各种类型验证码识别准确率
- 开发与维护标注工具
- 统一深度学习开发平台，自动化训练开发流程

**图像识别：**OCR 字符识别主要分为：字符切分、字符识别过程；字符识别模块，即为常规的图像识别过程。网络结构为多层卷积的网络结构。训练样本为180万张左右，主要是机器生成，采用多种字体，多种形式，保证样本的多样性。除去少部分切分错误，效果显著。

- 开发与优化字符切分，字符检测策略
- 采用机器生成大量样本，解决数据标注问题
- 多策略改进样本生成方案，提升数字，文字的识别准确率

**图像识别：**滑块验证码定位，基于opencv的图像边缘检测方式，获取滑块模版，窗口移动确定滑块位置，由前端完成随机轨迹滑动；

**文本分类：**交易文本数据分类，采用 TextCNN 网络模型，结合 TensorFlow 计算框架，准确率达到0.96

**文本分类：**大量杂乱的错误文案信息，需要统一错误类型，并返回给商户，优化商户体验，并减少人为工作量；准确率达95%

- 清洗数据，将训练数据转化为TF-IDF Vector，建立词袋
- 向量化的文本信息用于贝叶斯模型建模

**风控：**社交关系亲密度模型，基于400万用户的通讯录标签，结合运营商通话记录建模，输出用户与电话号码联系紧密程度，主要采用XGBoost算法；

- 结合通讯录备注数据，通话记录多维度构建相关特征，不断优化模型效果；
- 进一步完善数据报告，为贷后催收提供决策支撑

**风控：**基础数据库建设，提供多重关系的黑名单查询，六度空间，通讯录匹配计算，GPS定位等服务，为个人信用报告，反欺诈报告计算提供支撑。

### 数据获取(爬虫)

- 开发并维护多家机构网站的数据抓取
- 设计并开发配置化解析系统
- 模版化的设计，灵活的解析方式，实现html、xml、csv等多种数据格式的解析，极大提升开发效率，独自承担解析工作

**浙江中控技术股份有限公司 软件开发工程师**

2013.07-2014.06

主要负责windows软件开发

### 教育经历

2009-2013 浙江工业大学之江学院通信工程专业（本科）

### 爱好

喜欢编程和技术，不断阅读各种书籍，提升技术能力；

热爱运动，连续三年参加杭州国际马拉松，尤爱羽毛球，大学期间羽毛球校队成员；