

杜利强

(+86) 186-0019-8951 · keris.du@gmail.com · 算法工程师 · <https://keris.github.io/>

个人总结

目前在友信金服任职算法工程师。在此之前，曾在新浪微博担任搜索算法工程师，在诺禾致源负责开发短序列比对程序。在算法和工程方面拥有累计达 5 年的工作经验。拥有计算机科学硕士学位以及电子工程本科学位。对搜索引擎开发感兴趣，包括排序模型，点击率预估。此外，我还是一个机器学习和深度学习爱好者，业余时间完成优达机器学习工程师纳米学位，参加 kaggle 比赛。其他时间我会写一些文章，做一些感兴趣的个人项目。

教育背景

中国农业大学, 计算机科学与技术, 硕士研究生 2014.9 - 2017.1

GPA 3.8/4.0, 拥有两篇发明专利和一篇 ACM VRCAI 16 会议文章, 被邀做口头演讲。

中国科学技术大学, 电子科学与技术, 理学学士 2008.9 - 2012.6

GPA 3.3/4.0

技术能力

- 编程语言: C/C++, Python, Java, SQL, Shell
- 数据库: MySQL, MongoDB, Redis
- 框架: Hadoop, Spark, TensorFlow, PyTorch
- 开发工具: Visual Studio Code, Vim, Emacs, SVN, Git
- 操作系统: Linux/macOS

工作经历

友信科技 | Youxin Financial, 算法工程师 2018.10-至今

- 主导建立编码规范，在此基础上重构黑名单系统，使新规则添加友好、维护方便和监控高效。
- 负责迭代优化文本分类模型，对基于用户信息的搜索结果进行分类。
- 主导构建了一种通用的跟踪系统，使排查问题友好，并实时监控关心指标的变化。
- 基于用户关系网络设计和生成特征用于风险建模。
- 关键词: Hadoop, Hive, MongoDB, Redis, SVM, LR, Xgboost

新浪微博 | Weibo, 搜索算法工程师 2017.1-2018.9

- 负责微博搜索部门各业务线所用数据的接入、属性计算、分发和存储。
- 对已有的排序模型进行不断迭代改进站外文章的展现效果。
- 开发微服务服务于热门搜索和综合搜索，包括赞服务、转发服务和热门评论服务。
- 深度参与跨部门协作的微博故事搜索项目。
- 关键词: Hadoop, Hive, Kafka, Redis

北京诺禾致源生物信息技术有限公司 | Novogene, 生物信息开发工程师 2012.7-2013.10

- 开发短序列比对程序以检测 DNA 中的插入或删除。
- 性能优化: 利用 cuda 技术加速现有的比对程序。
- 关键词: SOAP, CUDA

项目

用户关系网络 | 友信科技 2018-至今

基于图数据库构建的用户关系网络，自然地建模不同用户之间的关系。关系网络使能更好的用户了解，支持用户关系直观探索以及生成特征用于风险建模。关系网络包含数十亿条边和数千万个节点。其中我负责：

- 性能调优：带来十分高效的数据注入至图数据库中，达数千条边每秒。
- 打造关系网络可视化工具支持用户关系直观探索。
- 设计和生成用户特征用于风险建模

关系网帮助拒绝了 0.8% 的欺诈交易，仍有十分巨大的潜力待挖掘。

关键词: JanusGraph, Gremlin, Hadoop, Spark, ElasticSearch, JavaScript, Python

黑名单系统 | 友信科技

2018-至今

黑名单系统作为一种机制以拒绝具有明显风险的申请人。它包含若干个表，保存了满足特定规则的交易涉及的用户信息，如银行卡、注册手机号码、紧急联系人，设备号等。黑名单提供 API 服务，在每次进件时调用。其中我负责：

- 开发并维护 API 服务。
- 重构系统以支持新名单添加以及旧规则淘汰友好。
- 日级别和周级别报表生成及推送。

黑名单帮助拒绝了 3% 至 4% 的欺诈交易。

关键词: Python, Redis, Tornado, MySQL, Kafka

网络筛查系统 | 友信科技

2018-至今

这个项目基于用户信息从百度搜索、贴吧、论坛等获取用户相关的搜索结果，然后对返回的文本结果进行分类，最后用于风险建模。其中我负责不断迭代改进文本分类模型。该系统帮助拒绝了 0.7% 的欺诈交易。

关键词: SVM, Python, Tornado, text-classification

优达机器学习工程师纳米学位

2018-2019

<https://graduation.udacity.com/confirm/G3M94KR7>

作为一个机器学习爱好者，我在业务时间完成了优达机器学习工程师纳米学位。

关键词: Machine Learning, Deep Learning, Reinforcement Learning, Python, TensorFlow

数据中间件 | 新浪微博

2017-2018

这个项目的定位是数据获取，数据属性计算，数据分发和数据持久化。它服务于微博搜索的各个业务线，包括热门搜索，综合搜索，微博故事和话题推荐。

关键词: C++, Python, Hadoop, Storm, Message Queue

文章搜索 | 新浪微博

2017-2018

这个项目通过包含长微博以及站点外文章以丰富微博搜索的结果。其中我负责不断迭代文章排序模型。此外，协同另一位同事改进和优化爬虫系统，确保内容的实时性。

关键词: C++, 排序模型

出版物

-
- Liqiang Du, Hong Chen, Shuli Mei, and Qing Wang. 2016. Real-time human action recognition using individual body part locations and local joints structure. In *Proceedings of the 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry - Volume 1* (VRCAI '16), Vol. 1. ACM, New York, NY, USA, 293-298. DOI: <https://doi.org/10.1145/3013971.3013974>