

施 华

求职意向：算法工程师

☎(+86) 15021408795 | ✉15021408795@163.com | 🏠 shihua.netlify.app | 📱 redblue0216



技能 & 爱好

- 数学能力-算法基础 基础研究方向为概率统计-理论计量
-算法方向 最熟悉的应用方向为时间序列、优化，工作培养的应用方向为计算机视觉和自然语言处理
- 工程能力-编程语言 Python,C,C++,Go,R,Matlab,Shell,SQL,Scala,Latex,Markdown,RestructuredText
-操作系统 Linux
-数据系统 MySQL,MongoDB,ElasticSearch,ClickHouse,InfluxDB,Sqlite3,Postgresql,Redis,MinIO
-科学计算 HDF5,Dask,Ray,MPI,Eigen,Numpy,Scipy,Pandas,Gonum,Gosl,Dataframe-go
-算法框架 Scikit-learn,Tensorflow,Deap,Networkx,Statsmodels,Pymc,BayesianOptimization,Gekko,Pulp,Pingouin,Sktime,Tslearn,Tsfresh,Mealpy,Linearmodels,Geatpy
-其他工具 RabbitMQ,Airflow,Celery,Flask,Consul,Zodb,Ansible,HDFS,Alluxio,Bootstrap,Spark,Kibana,Logstash,Docker,Kubernetes
- 外语能力-英语六级 一篇英文论文 Stochastic unit root model with generalized error distribution[J].Statistica Sinica(第一作者)
- 业余爱好-兴趣广泛 数学、编程、电子竞技、篮球、吉他

教育经历

西南民族大学

成都，四川

经济学院

2015年9月 - 2018年7月

- 数量经济学硕士，理论计量方向，平均分数：90.93/100
- 核心课程：高级微观经济学 I,II、高级宏观经济学 I,II、高等计量经济学 I,II、微观计量经济学、空间计量经济学、面板计量经济学、概率统计理论与方法、经济数学模型与 **Matlab** 应用、博弈论、概率理论与随机过程 I,II,III、高等数理统计方法 I,II,III、最优化理论 I,II、金融数学方法与应用、金融工程理论与应用、高等时间序列分析 I,II,III、金融时间序列分析 I,II,III、非参数计量经济学、计算统计 I,II、风险管理与数量建模、金融经济学、机器学习、统计学习、运筹学、实变函数 I,II、泛函分析 I,II、常微分方程 I,II、测度论 I,II、复变函数 I,II、数学物理方程 I,II

四川大学

成都，四川

数学学院

2014年9月 - 2017年7月

- 基础数学基地班旁听生
- 核心课程：数学分析、高等代数、实变函数、泛函分析、常微分方程、测度论、复变函数、数学物理方程、概率论基础、数理统计、多元统计、随机过程、抽象代数、解析几何、数值分析、最优化导论、矩阵计算、拓朴学、计算统计

中南财经政法大学

武汉，湖北

统计学院

2015年9月 - 2017年6月

- 高级计量经济学学术夏/冬令营
- 核心课程：高等数理统计、回归分析、时间序列分析、面板数据分析、微观计量导论、空间计量导论、统计模拟、金融计量专题

西南民族大学

成都，四川

经济学院

2010年9月 - 2014年7月

- 经济学学士，金融学专业，平均绩点：3.65/4
- 核心课程：数学分析、高等代数、经济数学基础、微观经济学、宏观经济学、商业银行管理、金融学、金融市场学、国际金融、计量经济学、证券投资分析、金融工程、公司金融、初级会计、财务报表与 **EXCEL** 建模、统计学、**C++** 金融实验、**R** 应用、投资学、财务会计、金融数学、金融衍生品、固定收益证券、风险管理、金融计量

平安科技

深圳，中国

- 量化金融晨星计划
- 核心课程: 数量入门、Python入门、金融入门、量化投资基础、量化分析风险与收益、股票和现金、数据分析与机器学习、债券与评级、金融数学进阶、金融编程 C++

工作经历

算法工程师, 绩元数据科技有限公司

2022年7月 - 2022年12月

- 岗位: 隶属数据研发部, 职级高级工程师(高级算法工程师)
- 职责: 主要负责绩元算法中心的设计、开发与维护(工程方面); 功率预测相关算法的优化与维护, 电力交易相关算法的设计、开发与维护(算法方面)
- 项目: 在职期间自主设计并开发的算法工具有: 绩元算法中心; 自主设计并开发的算法应用有电力交易算法模块和风电功率预测算法模块

算法工程师, 平衡机器有限公司

2020年3月 - 2022年6月

- 岗位: 隶属算法组, 职级 L4(算法组组长)
- 职责: 主要负责算法全生命周期工具套装 SEED 的设计与开发(工程方面)以及综合能源管理、低碳管理等主要业务相关算法的设计与开发(算法方面)
- 项目: 在职期间自主设计并开发的工程化工具有: 算法全生命周期工具套装 SEED, 自主设计并开发的算法应用有: 综合能源管理系统 [Balengy(包括功率预测算法模块、负荷预测算法模块、虚拟电厂算法模块、电力交易算法模块、气象预报算法模块)], 低碳管理系统 [NeutralC(包括交通流模拟算法模块、碳排放计算模块)], 其他非重点项目(包括用电客户画像系统、BI系统、气象减排系统)

算法工程师, 平安科技有限公司

2017年9月 - 2020年2月

- 岗位: 隶属宏观经济研究院开发五部算法组, 职级 C 类 2.3(中级算法工程师)
- 职责: 主要负责时间序列类统计算法的设计、开发与维护(算法方面), 量化风控系统基础算法库的设计、开发与维护(算法方面), 其他业务(包括因子自动挖掘系统、新闻舆情系统)相关算法的开发与维护(算法方面)
- 项目: 在职期间自主设计并开发的算法应用有: 股票风控评分系统 [Scoop(包括宏观情绪指数评分算法模块、微观事件评分算法模块、微观量价评分算法模块、宏观周期指数评分算法模块)], 新闻舆情系统 [RedScraps(包括主题抽取算法模块、主题公司关联算法模块)], 因子自动挖掘系统 [AlphaFinder(包括底层 DAG 运行调度工程模块 Cuber、因子自动组合选择算法模块)]

项目经验

算法全生命周期工具套装 SEED(平衡机器), 负责人-设计、编码与维护

2020年3月 - 2022年6月

- 项目简介: 该项目主要是为算法工程化提供一系列工具, 利用这些可快速构建一个轻量级的算法中心, 支持复杂算法应用。相关子模块有:
- 算法数据交互工具 Armor: 主要功能解耦算法与数据, 建立数据通道, 保障数据安全性与传输速率; 主要技术采取 RabbitMQ, Flask, gunicorn; 主要设计采用消息中间件进行数据交互, 并将数据读写服务化。
- 算法开发平台 Virtue: 主要功能提供算法应用的微服务管理和配置管理; 主要技术采取 Consul; 主要设计采取微服务 API 网关模式。
- 算法资源管理平台 Nadleeh: 主要功能进行资源隔离、调度和管理, 进程管理与应用监控; 主要技术采取 Kubernetes, Cesi, Supervisor, Prometheus; 主要技术采取容器边车模式。
- 算法统一回放平台 Seraphim: 主要功能在算法应用出现问题时, 提供数据回算和沙箱模拟功能; 主要技术有 Hbase, ElasticSearch; 主要技术采用冷热数据分离模式。
- 算法库平台 Veda: 作为算法平台 SEED 的基石, 主要功能提供算法包的存储、分发部署、环境配置和元信息管理, 支持云端和本地同步; 主要技术有 Sqlite3, Tar, FileSystem, MinIO, EnvironmentModules, Ansible; 主要设计采用观察这推拉模式、仿制 wheel 自研 vedapkg 打包格式、分布式环境配置操作。
- 算法运行时工具 Throne: 主要功能提供算法全生命周期中最重要的运行时工具, 包括针对单个算法点管理的算法模式运行时工具 Exia[主要功能有模型管理、参数管理、模型日志; 主要技术有 MinIO, MongoDB, ElasticSearch, Kibana; 主要设计采用对象池技术、DAG 技术和元编程技术实现主体框架(此处元编程技术包括描述符协议、上下文管理协议、单例模式、命令模式、工厂模式、eval 技术); 使用构建模式、组合模式与 MySQL 实现数据获取模块 DataAPI, 使用命令模式、MinIO、MongoDB 实现模型管理模块 ModelLibrary, 使用装饰器模式、ElasticSearch、Kibana 实现算法记录模块 LogDecoratorEK, 使用观察者模式、有限状态机实现模型监控模块 ModelMonitoring, 使用工厂模式、命令模式、Consul、SSH、Airflow 实现服务管理模块 ServerManager], 针对整个算法链条管理的算法应用运行时工具 Kyrios[主要功能有算法编排和工作流管理; 主要技术有 Hydra, Jinja2, Airflow, 主要设计使用控制、基础、扩展三大模块实现信息、结构、执行、状态、扩展和接口六大功能], 针对算法点辅助处理的算法特征管理工具 Dynames[主要功能特征处理; 主要技术 Feast; 主要设计采用目录服务]。

综合能源管理平台 **Balengy**(平衡机器), 负责人-设计、编码与维护

2020年3月 - 2022年6月

- 项目简介: 该项目实时采集产能和用能数据, 借助功率预测、气象数值预报、需求响应、电力交易等精准数据服务, 辅助用户提升能源管理水平, 优化能源交易策略, 完善虚拟电厂功能。主要算法模块有:
- 气象数值预报算法模块: 核心算法点有数值预报预处理 (使用 **Pygrib** 和 **HDF5** 封装程序包 **Meoteofence** 实现), 气象数据融合 (使用 **BayesianModelAverage** 实现), 风速修正 (使用状态空间卡尔曼滤波实现); 辅助算法工程点有矩阵计算缓存 (使用 **HDF5** 封装 **Raiser** 程序包实现)。
- 功率预测算法模块: 核心算法点有气象状况划分 (使用 **KmeansMiniBatch** 实现特征聚类), 功率预测 (使用 **BPNN** 实现回归预测), 提高泛化精度 (使用 **PSO-Bagging** 实现动态权重集成学习); 核心算法工程点有集成学习分布式框架 (使用 **Ray,Networkx,Kahn** 开发 **Cuber** 程序包实现); 辅助算法点有风电数据异常识别 (使用 **DBSCAN** 实现密度聚类), 光伏时序晨间突变抑制 (使用 **EMD** 分解出波动残差限值实现抑制), 光伏时序峰值附近抖动抑制 (使用差分限值和滑动平均实现限值低通滤波)。
- 负荷预测算法模块: 核心算法点有相似日划分 (使用 **GMM** 实现密度聚类), 负荷预测 (使用 **XGBoost** 实现回归预测); 辅助算法点有电网拓扑结构动态变化捕捉 (使用 **DAG** 连通结构实现计算量化), 专业领域特征工程。
- 电力交易算法模块: 核心算法点有电价预测 (使用 **GBR** 实现点预测, 使用分位数回归和核密度估计实现概率预测), 功率上报调整 (使用 **Qlearning** 实现功率分配的随机优化)。
- 虚拟电厂算法模块: 核心算法点有微电厂调度 (使用统计模拟和 **GA** 实现随机层次优化)。

城市低碳生态管理平台 **NeutraC**(平衡机器), 成员-设计、编码与维护

2020年3月 - 2021年3月

- 项目简介: 该项目基于云计算、大数据、区块链、机器学习、遥感分析等创新技术, 提供数字化城市低碳管理的闭环服务。主要算法模块有:
- 交通流模拟算法模块: 核心算法点有形成分布模拟 (使用泊松过程实现顶点对生成), 驾驶偏好优化 (使用线性规划实现边的权重生成), 最短路径选择 (使用 **Dijkstra,Fold** 实现路线选择), 驾驶行为模拟 (使用 **Gipps** 跟车模型实现交通仿真细化)。
- 碳排放计算算法模块: 核心算法点有碳排放计算 (使用 **IVE** 模型实现)。

股票风控评分系统 **Scoop**(平安科技), 负责人-设计、编码与维护

2018年6月 - 2019年12月

- 项目简介: 该项目从宏观、微观两方面对股票评分, 为量化择股提供参考, 构建股票风险池。主要的算法模块有:
- 宏观情绪指数模块评分算法模块: 核心算法点有混频与缺失值问题 (使用 **EM** 算法实现), 情绪指数构建 (使用三步 **OLS** 实现); 工程实现基于 **Numpy** 和 **Scipy** 自主研发。
- 宏观周期指数评分算法模块: 核心算法点有周期转换 (使用傅里叶变换实现频谱转换), 时序相似度量 (使用 **DTW** 实现不等长相似度量), 降维加速 (使用 **Kmeans-DTW** 实现时序聚类程序包 **KMDTW**); 工程实现基于 **Numpy,Scipy** 和 **Dtw** 自主研发。
- 微观事件评分算法模块: 核心算法点有多事件融合评分 (使用 **MDP** 模型实现主体评分模型, 使用经验法则分解 **MDP** 的转移概率为盈利、负债和状态转移三个概率, 使用频率计算盈利、负债概率, 使用 **Kmeans** 完成频率计算所需的状态标注分类, 使用基于贝叶斯思想的 **Beta** 分布来模拟状态转移概率, 使用单事件评分来估计 **Beta** 分布), 单事件行为评分 (使用 **OLS** 评估一般事件, 使用 **SUR** 评估公司关联类事件, 使用 **GARCH** 评估限幅的波动集聚类事件), 事件动态影响 (在 **Beta** 分布上加减单事件评分), 股票基础分 (根据财务数据建模得到), 最终评分 (基础分加上多事件融合分); 工程实现基于 **Numpy,Scipy,Statsmodels,Linearmodels,Scikit-learn** 自主研发。
- 微观量价评分算法模块: 核心算法点有股价预测 (使用 **GA-HMM** 实现平稳时间序列预测, 使用 **MS-AR** 实现非平稳时间序列预测), 量价评分 (根据涨幅评分); 工程实现基于 **Numpy,Scipy** 自主研发。
- 动态权重算法模块: 核心算法点有权重变动检验 (使用 **F** 统计量检验时间窗指定约束回归), 计算权重 (使用 **MLE** 估计受约束回归); 工程实现基于 **Numpy,Scipy** 自主研发。

因子自动挖掘系统 **AlphaFinder**(平安科技), 负责人-设计、编码与维护

2017年9月 - 2018年6月

- 项目简介: 该项目为多因子模型提供自动化因子筛选程序。主要算法模块有:
- **DAG** 运行调度框架 **Cuber**: 核心工程点有控制引擎 (使用 **Networkx** 实现), 计算引擎 (使用 **Ray** 实现分布式), 调度引擎 (使用 **Kahn** 算法实现并行化)。
- 因子自动选择算法模块: 核心算法点有因子适应度 (使用 **IC/IR** 或回归实现), 因子自动选择 (使用 **GA** 算法寻优)。

新闻舆情预警系统 **RedScraps**(平安科技), 成员-编码与维护

2018年6月 - 2020年3月

- 项目简介: 该项目根据热点新闻找出关联股票, 形成重点股票池, 为投资组合择股提供参考。主要算法模块有:
- 主题抽取算法模块: 核心算法点有主题抽取 (使用 **gensim** 实现 **LDA**); 核心工程点有文本存储 (使用 **MongoDB** 实现)。
- 主题公司关联算法模块: 核心算法点有主题关联 (使用 **Copulas** 实现 **Copula**)。

用电客户画像系统 (平衡机器), 负责人-编码与维护

2020年5月 - 2020年6月

- 项目简介: 该项目为江苏国网科研项目, 数字化用电客户分析, 优化运营管理。核心算法点有静态评分 (使用 **GDBT,XGBoost,Adaboost** 实现回归), 评分自适应 (使用二元 **Logit** 实现), 规则抽取 (使用 **Apriori** 实现); 工程实现基于 **Flask,Bootstrap,Scikit-learn**。

公司管理 BI 系统 (平衡机器), 成员-编码与维护

2021 年 9 月 - 2021 年 12 月

- 项目简介: 该项目为公司人事、业务人员提供高效办公工具, 主要算法应用有 OCR(使用 *Tensorflow* 实现 *CTPN*, *Densenset* 和 *CTC*), 自动报告 (使用 *reportlab* 实现)。

成都气象减排系统 (平衡机器), 成员-编码、维护

2020 年 3 月 - 2020 年 4 月

- 该项目为成都气象局污染减排治理科研项目, 主要功能包括重大污染案例管理、污染过程分析、相似过程匹配、臭氧案例过程识别; 工程实现基于 *R* 的 *Shiny*。

个人开源

轻量级算法平台 *Drip*(个人开源), 负责人-设计、编码与维护

2022 年 9 月 - 2022 年 12 月

- 该项目为个人开源的轻量生产级算法平台, 精简于平衡机器 *SEED* 算法平台, 使用 *Python* 重构开发, 覆盖从算法实验、算法运行前管理、算法运行时管理到算法运行后管理的全生命周期。相关子模块有:
- 信息管理 *Fiche*: *Fiche* 是一个元数据信息管理工具, 主要功能提供元数据信息的管理, 包括算法信息、模型信息、参数信息、应用信息和数据信息五大类信息; 主要设计采用将信息卡片化, 存储为一个个 *JSON* 字符串文档, 采用主从推送模型实现观察者模式, 便于信息中心化; 主要技术采用 *MongoDB* 作为后端信息数据的中心存储, 基于 *FastAPI* 实现的 *http* 服务使得该工具提供 *Python-SDK* 的同时也与语言无关。
- 存储管理 *Cask*: *Cask* 是一个存储管理工具, 主要功能包括项目本地打包, 远端存储管理两大功能; 主要设计仿制 *wheel* 文件格式, 远端存储采用 *MinIO* 技术。
- 调度管理 *Liquid*: *Liquid* 是一个 *Pipeline* 工具, 主要功能提供算法应用快捷组织功能; 主要技术使用基于 *Pluggy* 的 *Hook* 技术。
- 数据通道 *DaShare*: *DaShare* 是一个数据服务接口包, 主要功能是为底层数据向外提供了一个接口服务, 可实现用户权限验证和流量限制。*DaShare* 采用 *Client-Server* 架构设计, 服务端使用 *HTTP* 协议构建数据服务的 *OpenAPI*, 暂时采用 *GET-method*, 客户端封装为 *Python-SDK*, 提供 *Python* 数据接口功能; 用户权限采用 *token* 机制。
- 特征管理 *Atom*: *Atom* 是一种特征管理工具, 以数据和算子作为基本概念, 数据为基础数据用于训练特征和构建特征; 算子为基于固定一个或多个数据集进行新特征生产的流程, 可以是一个简单直接计算函数, 也可以是一个复杂的算法模型, 还可以是算法模型和直接计算想结合的组合体。*Atom* 的特点是对由数据衍生的算子进行了数据关联、统一管理, 并直接提供了服务功能, 使得每个算子可以直接实现在线实时计算特征, 为主体算法模型服务, 提高模型精度。
- 日志管理 *Trajectory*: *Trajectory* 是一个日志管理工具, 主要功能提高日志统一集中管理功能, 主要技术基于 *Logging* 提供一个 *MongoDB* 的日志转发器, 包括日志过滤、日志格式化功能。

分布式集成学习框架 *Cuber*(个人开源), 负责人-设计、编码与维护

2022 年 9 月 - 2022 年 12 月

- 项目简介: *Cuber* 是一个集成学习框架, 主要用于各种集成学习算法的开发, 具有灵活动态可配置的特点。*Cuber* 主要由三大模块组成, 控制引擎、计算引擎和调度引擎。控制引擎依托于 *Networkx* 的有向无环图技术, 需要以计算引擎和调度引擎为基础, 运行时加载。计算引擎支持 *Ray, Dask*; 调度引擎支持 *Kahn* 算法。

学术论文

- 施华, 王艳琴. 债转股企业道德风险的演化博弈分析 [J]. 财经界 (学术版), 2017(02):104-105.
- 施华. 广义误差分布下的随机单位根模型. 优秀硕士毕业论文.
- 施华, 毛瑞华. *Stochastic unit root model with generalized error distribution*[J]. *Statistica Sinica*.