

个人简历

个人资料

姓名: 王枫

性别: 男

学历: 全日制硕士研究生 (数学系应用统计方向)

出生年月: 1988 年 02 月

祖籍: 山东曲阜

户口: 北京市

手机 (微信): 18810450335

E-mail: science_maths@qq.com

个人网站: <https://mapleureka.github.io/index-ch.html>

工作年限: 8 年 (2013 年 7 月至今)

求职意向: 软件工程师。Java Web 全栈+混合移动 app、NLP 与数据挖掘、算法开发。个人 web 开发技术栈为 Spring Boot、Spring Cloud; React、EasyUI、Material-UI、Bootstrap; Cordova。

学术研究方向 (业余): 基础数学与理论物理

注: 本人目前在职, 提前看一下外部机会, 如果合适的话, 一个月内可以入职。工作单位可以在北京郊区, 避免拥堵。



技能评价

理论技能:

掌握系统的数学、统计学、物理学与软件工程知识, 对常用的数据分析挖掘算法、NLP 算法、图像视频处理算法、频谱分析有所研究。

编程语言:

C/C++、Java、R/Julia/Matlab、Shell/Bash、HTML/CSS/JavaScript、TeX/LaTeX

软件工程技能:

- 数据挖掘与科学计算:** 精通 R/Julia 数据挖掘分析技术; 精通 C/C++/Java/Matlab/Julia/R 科学计算编程; 精通音频、图像视频 (FFmpeg、OpenCV)、自然语言等相关开发;
- web 前端:** 熟练使用 Bootstrap 框架、Material-UI/EasyUI 框架、React 相关库; 熟练使用数据可视化框架 Highcharts; 掌握前后端交互的 AJAX 技术;
- web 后端:** 熟悉 Spring Cloud 微服务开发, 了解基于 Vert.x 与 Spring WebFlux 的响应式微服务架构; 熟练使用 Vert.x 全家桶开发, 精通 Spring Boot 全家桶开发; 精通 Spring Security 以及 Apache Shiro 安全认证框架; 精通单点登录 CAS 框架使用;
- 手机移动端:** 了解 web app、native app 与 hybrid app 相关开发, 熟练使用 Cordova;

- 5、**数据库以及相关存储**：熟练使用常见数据库 MySQL、Oracle、PostgreSQL、Hbase、Hive；熟练使用 ElasticSearch；
- 6、**数据开发**：熟练使用 Kettle、Oracle SQL developer、Oracle GoldenGate、Flume；熟练使用消息中间件 RabbitMQ 与 Kafka；精通大数据流计算框架 Spark Streaming/Flink 以及数据实时计算与落地；熟练使用 Hbase、Hive/Kudu/Impala、Spark；熟练使用大数据平台上的 Kerberos 认证工具；
- 7、**嵌入式开发**：精通跨平台的 C/C++ 软件开发，熟练使用 GNU Build Chain，熟练使用集成开发工具 Anjuta 与 Qt；了解 C51 单片机、ARM、STM32 以及 RTOS。
- 8、**运维**：精通 Linux 与 Windows 服务器运维、docker 运维以及大数据平台使用与运维；

个人特质

思维活跃、专业知识面广、具有很强的接受和学习新知识的能力，做事认真负责，善与人交流、有团队意识，为人积极乐观。

工作经验

2013.07-2016.04 北京神州泰岳软件股份有限公司 NLP 与数据挖掘工程师

部门：研发中心-大数据部

工作内容：

2013.07-2013.10 开发团队小组的大数据挖掘分析平台运维

软件环境：CentOS, Oracle, Cloudera 大数据平台

开发工具：Shell/Bash

内容介绍：

- Linux 装系统、磁盘做 RAID、LVM 管理、用户管理以及大数据平台搭建。

责任描述：Linux 服务器运维。首次接触大数据，学习大数据各个组件的搭建、维护以及使用。

2013.10-2014.06 公司日常数据挖掘分析任务

软件环境：CentOS, Oracle, Cloudera 大数据平台

开发工具：R 语言

算法工具：回归分析，多元统计分析，贝叶斯统计分析，时间序列分析

内容介绍：

- 业务数据的常用统计分析：频率分布直方图，概率分布拟合，分布的峰度系数与偏度系数计算，箱线图等；
- 共线性诊断以及岭回归分析、Lasso 回归；
- 3 西格玛准则的静态基线与动态基线；
- 异常值处理分析、聚类分析、判别分析、因子分析；
- 序列模式挖掘、关联规则挖掘、协同过滤推荐；
- 基于矩阵 SVD 分解的浅层语义关联分析。

责任描述：公司日常业务数据的数据挖掘分析并提供分析报告，以及撰写挖掘分析算法文档。

2014.06-2015.03 公司大数据平台 NLP 与文本挖掘——Cloud UP 产品线

软件环境：RHEL , Oracle

开发工具：Shell/Bash, R 语言, Java, Spark, Spark Streaming

算法工具：隐马尔科夫链，条件随机场，贝叶斯分类器，Kmeans, LDA 主题模型，协同过滤推荐算法

内容介绍：

- 基于 HMM 模型的中文分词模块开发，并提供 Java 接口；
- 新词发现模块开发，并提供 Java 接口；
- 基于 CRF 模型的命名实体识别模块开发（主要提供地名提取以及手机品牌提取），并提供 Java 接口；
- 短语结构句法分析以及依存结构句法分析模块开发，并提供 Java 接口；
- 文档向量化、特征抽取、文档聚类分类、推荐算法模块开发，并提供 Java 接口；
- Spark 平台运行接口程序处理海量文本数据。

责任描述：全部由个人独立完成

2015.03-2015.08 民生银行服务器监控项目

软件环境：RHEL, Oracle

开发工具：Shell/Bash, ODBC/JDBC, R 语言, Java

算法工具：局部动态回归，潜周期时间序列模型，局部关联规则，动态基线

项目介绍：

- 使用 R 语言从 Oracle 提取数据，清洗数据转化数据；
- 在 R 语言中使用正则表达式在日志文件中提取相关数据转化为 Data Frame；
- 使用 R 语言对内存使用率使用局部动态回归以及潜周期时间序列模式进行预测，使用 Java 调用 R 语言提供 Web Service 服务供前端工程师调用；
- 对服务器 CPU 使用率，内存使用率，温度等指标使用 R 语言计算静态基线与动态基线，使用 Java 调用 R 语言提供 Web Service 服务。

责任描述：除了一些数据提取的复杂 SQL 由 DBA 完成外，其余都是由个人独立完成

2015.08-2015.10 中信证券大数据挖掘分析平台项目

软件环境: RHEL 7, Cloudera 大数据生态圈

开发工具: Shell/Bash, R 语言, Java, Presto, ODBC/JDBC, Spark

项目介绍:

- 大数据平台搭建, R 语言数据挖掘平台搭建;
- 基于 ODBC/JDBC 的 R 语言与 Oracle、Hbase、Hive、Spark、Presto 的交互模块搭建;
- R 语言调用 Java 语言模块搭建、Java 语言调用 R 语言模块搭建以及 R 语言程序包开发。

责任描述: 主要负责数据挖掘平台搭建, 个人独立完成

2015.10-2016.01 车辆道路监控产品

软件环境: Windows

开发工具: C++, MATLAB, OpenCV

算法: 形态学处理算法, 高斯滤波器, 卷积神经网络

项目介绍:

- 图像预处理以及字符提取的 MATLAB 建模;
- 对提取字符进行卷积神经网络模型训练与预测;
- 将 MATLAB 建模工程转化为 C++实现。

责任描述: 负责车牌号识别模块接口的开发与迭代调优, 个人独立完成

2016.01-2016.04 物联网智慧线产品

硬件环境: STM32

软件环境: Linux

开发工具: R 语言, MATLAB, C++, libsvm, GNU build Chain

算法: 卡尔曼滤波器, 支持向量机, 短时傅里叶变换, 小波分析

项目介绍:

- R 语言处理分析电磁信号数据以及异常值识别算法建模;
- 对电磁信号进行频谱分析提取异常信号波动特征后使用 SVM 算法进行机器学习;
- 将 R 语言与 MATLAB 建模程序代码转化为 Linux C++实现。

责任描述: 负责项目异常值识别模块接口的开发与迭代调优, 与小组成员共同完成

2016.05-2017.04 中国科学院北京基因组研究所 生物信息学大数据处理分析工程师

部门: 精准基因组医学重点实验室

工作内容:

2016.05-2017.04 与高原适应遗传基础研究相关的国家自然科学基金项目 (精准基因组医学重点实验室)

硬件环境：天河二号超级计算机

软件环境：SLURM

开发工具：Shell/Bash, R 语言, Java, GNU build Chain

项目介绍：

- 写 shell 脚本调用基因数据拼接、比对、关联分析等软件处理二代测序仪测出的人类全基因组数据；
- 使用 grep、sed、awk 以及 R 语言进行数据抽取以及分析挖掘；
- 天河二号 SLURM 任务管理器的作业监控；
- 海量基因组数据的传输与保存。

责任描述：根据科研项目需求，设计数据处理分析流程，转化为可供论文发表的数据分析结果

2016.05-2017.04 科研项目组的日常工作（精准基因组医学重点实验室）

硬件环境：基因组研究所 PBS 计算机集群

软件环境：PBS

开发工具：Shell/Bash, C, C++, Java, R 语言, GNU build Chain

内容介绍：

- 写 shell 脚本调用基因数据拼接、比对、关联分析等软件处理二代测序仪测出的人类全基因组数据；
- 使用 grep、sed、awk 以及 R 语言进行数据抽取以及分析挖掘；
- PBS 任务管理器的作业监控；
- 海量基因组数据的传输与保存；
- 基因组数据分析软件开发；
- 高性能计算机装机以及日常维护。

责任描述：根据科研小组的教学与科研需求，指导学生软件工程相关知识以及生物信息学数据处理分析

2017.05-2018.04 软通动力技术有限公司 大数据开发工程师

部门：华北高科技事业部

工作内容：

2017.05-2018.04 铜仁凤凰机场海关智慧查验项目（同方威视人工智能研发中心-平台开发部）

软件环境：CentOS 7, Hadoop, Hbase, Flume, kafka, Neo4J, RabbitMQ, MySQL

开发工具：Java, Spark Streaming, MyBatis, Restful API

项目介绍：

- Ambari 大数据集群的搭建与运维；
- Flume 和 Kafka 采集数据（安检机的日志数据和图像数据）；

- 使用 Spark Streaming 分布式实时解析 JSON 字符串以及分布式数字图像处理;
- HBase 数据存储与图像的 HDFS 存储;
- Web 后台大数据分析任务管理器开发;
- Neo4j 同行数据分析模块的数据库存储与查询分析;
- 海关违禁品查验相关数据挖掘算法研究。

责任描述: 根据甲方业务需求, 自己独立完成自己负责的模块, 最后融合在一起

2018.05-2019.11 文思海辉技术有限公司 高级设计开发工程师

部门: 金融事业部

工作内容:

2018.05-2018.08 中国人寿反欺诈评分服务项目 (中国人寿研发中心三部六室-AI 小组)

软件环境: Hbase, Kerberos, WebLogic, CentOS 7

开发工具: Java, axis2, Spark, SoapUI

项目介绍:

- 对大数据集群 HBase 数据服务进行 Kerberos 认证;
- 利用 Spark 的随机森林模块进行训练模型, 然后对请求数据进行预测并返回欺诈值;
- WebService 服务端开发 (axis2 框架), 提供反欺诈接口服务; 将客户的每次请求数据与返回结果记录到 hbase 表, 将 hbase 映射为 hive 表做一些统计分析;
- Web Service 服务运维, 以及基于 docker 容器技术的自动化云部署。

责任描述: 根据甲方业务需求独立完成整个项目的开发、部署与运维

2018.09-2018.12 中国人寿标准体识别服务项目 (中国人寿研发中心三部六室-AI 小组)

软件环境: Hbase, Kerberos, WebLogic, CentOS 7

开发工具: Java, axis2, xgboost, SoapUI

项目介绍:

- 对大数据集群 HBase 数据服务进行 Kerberos 认证;
- 利用 xgboost 进行训练模型, 然后对请求数据进行预测并返回欺诈值;
- WebService 服务端开发 (axis2 框架), 提供标准体判别接口服务;
- Web Service 服务运维, 以及基于 docker 容器技术的自动化云部署。

责任描述: 根据甲方业务需求独立完成整个项目的开发、部署与运维

2019.01-2019.03 中国人民人寿保险审计开发项目 (中国人民人寿保险股份有限公司运维二处)

软件环境: Hive On Spark, Impala, Kerberos, Oracle, CentOS 7

开发工具: Beeline, Hue

项目介绍:

- 承保、理赔、退保审计风险点的 Hive SQL 开发与优化。

责任描述：根据甲方业务需求文档完成自己负责的审计风险点模块的开发与部署

2019.04-2019.06 中国人民人寿保险数据同步开发（中国人民人寿保险股份有限公司运维二处）

软件环境：Oracle, Oracle Golden Gate, Hive, Spark Streaming, Kafka, Kerberos

开发工具：Java, Spark Streaming, Kafka

项目介绍：

- OGG 数据采集软件部署；
- 大数据集群各个组件的 Kerberos 认证以及相关代码封装；
- 将 Oracle 的 DDL 与 DML 操作实时通过 OGG 转发到 Kafka 集群；
- 开发 Spark Streaming 程序消费 Kafka 集群各个 topic 中的数据，并将数据转换为 DataFrame 然后通过 Spark SQL 导入到 Hive。

责任描述：根据甲方业务需求独立完成整个项目的开发与部署

2019.06-2019.07 北京市科实五金有限责任公司 MES 系统（北京金隅集团股份有限公司）

软件环境：MySQL, Windows 10, GitLab

开发工具：Java, Spring boot, TkMyBatis, jQuery EasyUI

项目介绍：

- 对 MySQL 数据库的 CRUD 以及相关数据的页面表格展示功能实现；
- 下拉列表菜单、文本框检索、数据分页等功能实现；
- 图片存储、图片上传与下载功能实现。

责任描述：整个项目划分为若干模块，个人完成自己负责的上面列举的部分

2019.08-2019.11 贴源层数据同步开发（中国人民人寿保险股份有限公司开发运维二处）

软件环境：Oracle, Hive, Kudu, Impala, Oozie, Kerberos

开发工具：Java, Spring Cloud, Sqoop, Hue

项目介绍：

- 大数据微服务集群搭建以及微服务开发；
- 使用 Impala 在 Hive 与 Kudu 之间相互导数；
- Impala JDBC 的 Kerberos 认证代码封装；
- 开发抽数功能模块，实现 Impala JDBC 从 Hive 表取数以及写入本地文件；
- 写 shell 脚本调用 Sqoop 从 oracle 往 hive 导数，并配置 Oozie 定时任务；
- 数据同步时耗统计分析汇总功能开发。

责任描述：整个项目划分为若干模块，个人完成自己负责的上面列举的部分

2019.12-至今 **股份有限公司**

软件工程师

部门：医保数据技术与服务事业部

工作内容：

2019.12-2020.02 医保数据技术开发 (***项目)**

软件环境：MySQL, Hive, Kudu, Impala

开发工具：R 语言, Java, Spring Boot

算法工具：最小二乘法、多层感知机、RBF 网络

项目介绍：

- 使用局部最小二乘法对青岛市社保人员流入流出预测；
- 使用 Rprop 算法对多层感知机回归网络进行模型训练，并对青岛市社保人员流入流出预测；
- 使用径向基函数网络建模，并对青岛市社保人员流入流出预测；
- 将预测模型封装到 R 语言脚本，并设定定时任务将预测结果写入 MySQL；
- 使用 Spring Boot 开发模型预测的 web 服务供前端调用。

责任描述：整个项目划分为若干模块，个人完成自己负责的上面列举的部分

2020.02-2020.03 医保数据技术开发 (***项目)**

软件环境：MySQL

项目介绍：

- MySQL 存储过程开发：超住院时间预警，通过对参保人的住院时间进行分析，对于某病人住院时间超过该病种平均住院时间的情况进行预警；
- MySQL 存储过程开发：死亡再住院预警，对病人住院以首次死亡状态出院后再入院进行预警，筛选出部分预警数据供前台页面展示；
- MySQL 存储过程开发：偏差病例预警，对病人住院以首次死亡状态出院后再入院进行预警，筛选出部分预警数据供前台页面展示。

责任描述：整个项目划分为若干模块，个人完成自己负责的上面列举的部分

2020.03-2020.05 上海市多家医院数据开发支持

软件环境：MySQL, Oracle, Hive

开发工具：Java, SQL

项目介绍：

- Hive 与 Oracle 之间导数，Oracle 数据更改等；
- 数据处理与数据标准化。

责任描述：整个项目划分为若干模块，个人完成自己负责的上面列举的部分

2020.05-2020.11 健康险人员管理系统 (***项目)**

软件环境: MySQL, Redis, vsftpd, nginx, Git

开发工具: Spring Boot, Spring Cloud, Vue

项目介绍:

- MySQL、Redis、vsftpd、nginx 的部署;
- 前后端的配置与编译打包;
- web 前后端分离的负载均衡模式部署;
- 移动前端的扫码登录调试与 bug 修复;
- 移动前端嵌入到云助理 app 频道 (类似于微信订阅号) 与微应用的相关测试与 bug 修复。

责任描述: 整个项目划分为若干模块, 个人完成项目的测试与 bug 修复以及产品部署上线。

2020.11-2020.12 大数据高净值个人居民筛选 CRS 项目 (*****项目)

软件环境: Cloudera 大数据集群, CentOS 操作系统, SQL Server, Oracle

开发工具: shell/bash, beeline, impala, Sqoop

项目介绍:

- 抽取所有省份的 ods 保单到大数据平台;
- 抽取所有省份的精算数据 (包括股份精算数据和集团精算数据) 到大数据平台;
- 保单数据加工 (包括数据关联、排序、筛选等), 进行高净值个人居民筛选;
- 将筛选结果抽取到 Oracle 数据库。

责任描述: 个人完成改项目在生产环境上全部操作。

2021.01-至今 大病银保信保送 (*****项目)

开发工具: Spring Boot, MyBatis, EasyUI, MySQL

项目介绍:

- 搭建 web 系统框架 Spring Boot+MyBatis+EasyUI;
- Excel 文件上传界面, Excel 上传到 web 后端、web 后端解析 Excel 内容并写入数据库;
- web 前端文本输入框的规则校验。

责任描述: 个人完成上面列举的全部功能。



扫码关注个人微信公众号