

☎ 13701890718
✉ johnnyrxia4@126.com
🌐 <https://github.com/jamesxia4>
📡 <http://blog.jamesxia4.me>
🏠 浙江大学玉泉校区25舍220室

夏元昉

商业智能研究员/数据仓库工程师/数据挖掘工程师

项目与工作经验

EXPERIENCE

滴滴出行大数据研究院-代驾数仓 2016.01 ~ 至今

数据仓库开发 (实习)

- 使用 **Hive/Hadoop/Perl/Python** 完成代驾数仓开发
- 参与代驾自营司机客户端日报开发, 使用**动态分区**统计客户端每日累计在线时长计算日薪
- 独立完成**乘客画像**一期, 了解画像类开发常用指标, 解决了**数据倾斜**问题
- 独立完成滴滴快的代金券**数仓对接**与券报表开发, 使用 **grouping sets/rollup/cube** 支持多维多层次分析
- 独立完成"暗夜"**网安数仓**开发, 使用**开窗函数**实现了订单抽样
- 独立完成**客服门户**一期, 使用 **Lateral View/UDTF** 分层分析投诉原因
- 参与快的**遗留代码**向滴滴集群的迁移工作, 独立完成部分代码的**性能优化**, 使用**减少子查询/select */先限分区再限字段/先取数后计算/JVM重用/参数调整**等方法解决迁移后超时的问题
- 编写**Python**脚本, 使用**正则**替换的方法实现了快的代码的批量迁移
- 日常负责**核心KPI**报表、**司机画像**、**城市运力热力图**、快的 **UDF** 的维护
- 参与数据质量检查与故障定位
- 参与上游**需求对接**与下游后台开发**联调**

网易游戏GameUX-CoolData研究组 2015.08 ~ 2016.01

数据工具开发 (实习)

- 调研竞品**腾讯 WeTest** 舆情监控系统语料标注模块、**数据堂**众标平台
- 使用 **Sphinx/Markdown/MindManager** 完善设计文档、使用 **Axure** 独立完成原型开发、与UI配合迭代交互稿、完成数据库设计
- 基于 **Spring** 独立完成后端开发, 着重了解了 **Spring DAO**, 使用 **JUnit/JMock** 完成测试
- 根据交互稿基于 **jQuery** 独立完成前端开发, 手写了**进度条**、**卡片答题框**等部件
- 参与代码集成、联调、迭代、前后端**性能优化**, 根据日志定位线上Bug
- 使用**Python**对已爬取的游戏评论抽样生成标注数据集, 生成算法评价

个人项目

知识图谱-Slot Filling/Entity Linking

- 使用 **SVM** 与规则完成了维基百科Infobox **Slot Filling**算法
- 使用**朴素贝叶斯**实现了**Entity Linking**算法, 加入实体类型判断提升主题一致性, TAC 2015 EDL第二名

教育经历

EDUCATION

2014 ~ 至今

浙江大学

- 计算机科学与技术 (硕士)
- TAC 2015 EDL Track 第二名

2010 ~ 2014

上海交通大学

- 计算机科学与技术 (本科)
- MCM/ICM 2013 美国大学生数学建模竞赛 二等奖

技能

SKILLS

商业智能/数据仓库/数据挖掘

Hive/Hadoop/Perl/Python

- 熟悉 **Hive QL**, 会写复杂的查询, 了解 **HQL基础** 与**进阶** 知识, 了解 **UDF** 开发; 会用 **Sqoop**, 了解 **Hadoop** 与 **HDFS**
- 了解常见的 **HQL优化** 技巧并能在项目中加以实践, 会排查常见数据质量故障
- 熟练运用 **Perl / Python** 开发ETL脚本, 会用**Python**写**爬虫**, 会用**正则**
- 熟悉 **Bash**, 能开发自动化脚本, 熟练运用 **Linux** 命令

Web 开发

Javascript/CSS/HTML

- 熟练使用 **jQuery/ECharts/D3.js** 完成开发, 会用 **Chrome DevTools** 调试
- 熟练使用 **Bootstrap** 完成简单的页面布局, 能按交互稿完成页面的开发
- 使用 **Jade/Gulp** 生成了**这份简历** (见**Github**)

Java/Python

- 熟悉 **Java/Python**, 会用 **Spring/Django** 搭建后端及提供 **RESTful API**, 会用 **JUnit** 编写测试

其他

- 有使用 **MySQL**、**Apache**反向代理的经验
- 日常使用 **Git** 与 **Vim** / **IntelliJ** / **WebStorm**, 也乐于尝试新工具

其他

- ★ CET-4:**651**, CET-6:**619**, 听说读写优秀