贺佳玮

• **职位意向**:数据挖掘应用、机器学习相关、后台开发、全栈工程师

• Email: jayveehe@gmail.com

• GitHub: https://github.com/JayveeHe

• Blog: http://jayveehe.github.io

技术栈

• 机器学习:

- 。 硕士期间学习研究NLP, 主攻文本挖掘与分析方向。
- 。 硕士论文课题方向为——商品评论的有用性预测及重要性排序。提出一种基于LDA的写作风格表示方法。
- 。 学术论文: ICKEA 2016 《A Supervised Method for Ranking Reviews Based on Latent Structure Features》
- 。 掌握常用的机器学习算法(LR、NB、SVM、RF、HMM、CRF等),了解机器学习前沿技术(神经网络、词向量等)。

• Python:

- 。 熟悉, 作为个人项目主力语言。
- 。 常用Web框架为 Flask 。
- 熟练使用 Scikit-Learn 、 gensim 、 pandas 等第三方机器学习及数据分析的工具库。

• Java:

- 。 掌握,有基于OOP思想的项目开发经验,能独立设计开发整个项目。
- 。 掌握Java常用特性, 了解JVM内存原理。

数据库:

。 NoSQL: 掌握,常用的非关系型数据库为 MongoDB ,使用过 Redis 做为缓存。

。 SQL: 了解基本SQL语法,常用关系型数据库为 MySQL 。

• Linux操作:

。 日常使用 os x 进行开发,掌握Unix-like系统的基本命令。

• Android:

。 了解, 能独立开发简单的APP。

• 前端:

- 。 掌握基本的原生 JavaScript 语法,能编写函数进行基本的HTML元素控制与相关的数据交互操作。
- 。常用 jQuery 、 Bootstrap 等框架,接触过 AngularJS 。能独立设计简单的网页。
- 。 熟悉爬虫、抓包的相关流程。

• Big Data生态:

- 。 了解基本的 Spark 数据处理方法
- 。 了解Map-Reduce模型

教育背景

- 北京邮电大学 研究生 电子与通信工程专业 自然语言处理方向 (2014.09 2017.04)
 - 。 语言水平: 英语六级, 无障碍阅读外文文献与技术文档。
 - 专业基础:研究掌握了本方向的理论知识,包括数据挖掘算法与自然语言处理的相关理论。
- 北京邮电大学 本科生 信息工程专业 (2010.09 2014.06)
 - 。 专业基础:学习并完成计算机相关学科课程,包括计算机网络、数据结构、数字逻辑、Java程序设计、C程序设计等。

实习经历

- 深圳云天励飞技术有限公司 【算法工程师】 (2016.07 2016.09)
 - 1. 负责计算机视觉相关的算法原型实现。
 - 2. 利用高维LBP特征进行人脸性别与年龄识别,性别识别准确率97%以上,年龄识别绝对差6岁左右。
 - 3. 研究实现高维特征的稀疏化映射矩阵,以实现特征降维,并主要研究了涉及 L1 正则项的最优化方法。 实现模型大小压缩 100 倍,准确率仅下降 1~2%。
- 北京拉布科技有限公司 【算法工程师】 (2015.06 2016.02)
 - 1. 负责情境识别算法的设计与实现, 利用CRF模型对用户状态进行情境标注。
 - 2. 作为负责人参与微信公众号文章分析(用户阅读兴趣画像)项目,设计技术方案,并与团队合作完成产品,同时负责与客户对接以进行产品迭代。
 - 3. 采用微服务架构,基于Python的Flask框架向其他模块提供RESTful API。
 - 4. 针对不同业务要求,设计MongoDB的存储数据结构,并建立合适的索引优化查询。
 - 5. 综合使用Python、Node.js、JavaScript和相关的jQuery、Bootstrap框架设计实现公司的内部Web

- 工具,提高了数据处理的效率。
- 6. 负责姿态识别算法模块的Android移植,并提供demo供以测试。
- 7. 掌握基本的Linux部署操作命令和基于Jenkins的CI自动化部署流程。

项目经历

- <u>MusicTaster</u> — 种Word2Vec转Song2Vec、Aritst2Vec的实践 【个人项目】 *(2016.12-2016.03)*
 - 1. 基于爬取的网易云音乐200w+歌曲、20w+歌单、7w+歌手等信息,训练Song2Vec模型。
 - 2. 实现任意歌单内歌曲的聚类及前端可视化。
 - 3. 实现曲库内(20W+歌曲)任意多首歌曲的相似度搜索,并返回最相似的Top K歌曲。
 - 4. 提供一套更丰富的API
- **Senz情境感知**——用户情境识别SDK 【团队项目】 (2015.06 2016.02)
 - 1. 嵌入Senz SDK 的移动应用可获得用户级别的情境识别能力,包括用户画像(职业、性别、领域、喜好等)、姿态识别(静坐、步行、跑步、乘车)、场景识别(在家、在公司、在上班路上、在回家路上、在电影院、在商圈等)。
 - 2. 负责基于手机应用列表的用户画像算法的设计实现,为高阶用户画像提供基本信息。
 - 3. 负责用户场景识别算法的设计实现,定时批量标注用户所处场景(30w/天,场景事件识别准确率 60%+)。
- <u>PageTemper</u>——微信公众号用户兴趣画像 【团队项目】 (2015.08 2016.02)
 - 1. 利用微信公众平台的API,配合后端相关服务,记录公众号粉丝的阅读情况(用户级),并分析出 具体用户的阅读兴趣,为接入的公众号提供用户阅读兴趣数据及相应的统计信息。
 - 2. 负责整体技术方案设计、微信公众平台API研究、文章分析算法模块的设计与实现。
 - 3. 基于Flask提供文章分析API,数据存储使用MongoDB。
- Shooooty——Chrome浏览器弹幕扩展程序 【合作开发】 (2015.10 2015.12)
 - 1. 基于Flask框架与Leancloud云引擎部署后端程序、RESTful API实现前后端数据交互。
 - 2. 研究Chrome Extension API,并结合jQuery对网页的HTML元素进行操作,实现在任意网页发送/接收弹幕的功能。
- <u>淘宝买不买</u>——淘宝、天猫商品评价聚类 【独立开发】 (2015.03 2015.04)
 - 1. 给定一个淘宝&天猫商品页的URL,爬取该商品的评价,对评价中的关键字、句子组成进行聚类分析并呈现. 得出各组关键词下的具体评价内容. 供网购者参考。
 - 2. 部署于阿里云引擎,使用Java开发,涉及Servlet、HTTP爬虫技术、文本处理技术、数据挖掘算法等。
- 第三方新浪微博Android客户端 【独立开发】 (2014.10 2014.12)
 - 1. 使用新浪微博开放平台SDK,实现的功能有: 查看登陆用户所有微博, 查看登录用户所关注用户的

最新微博,查看评论,针对某条评论进行回复,评论、转发某条微博。

- YGameFrame — 自制2D Android游戏引擎 【团队项目】 (2014.06 2014.10)
 - 1. 基于jBox2d和OpenGL自制2D游戏引擎, MVC架构。
 - 2. 负责引擎地图模块的方案实现、游戏demo的动作逻辑设计。

活动经历

- 阿里巴巴**2016**年天池数据竞赛——<u>阿里音乐流行趋势预测大赛</u> (2016.05 2016.06)
 - 1. 根据样本用户的历史播放数据,预测未来两个月内每天各歌手的被收听量。
 - 2. 作为团队(3人)主要参与者,进行了包括数据清洗、特征分析、正负样本选取、模型建立与算法 实现等工作,最终取得第一赛季139名的成绩(139/5476)。
- 驾数据 领未来 <u>大众汽车数据创新大赛</u> (Hackathon) (2015.07 2015.07)
 - 1. 作为4人团队的一员参与竞赛并最终获得最佳商业前景奖(头等奖)
 - 2. 负责汽车原始GPS轨迹数据的清洗与聚类分析,并结合百度地图的JS SDK制作了轨迹可视化 demo,供以演示最终的算法结果。
- 阿里巴巴**2015**年天池数据竞赛——<u>阿里移动推荐算法竞赛</u> (2015.05 2015.07)
 - 1. 根据用户对商品集一个月的历史浏览、收藏、添加购物车、购买记录,推荐、预测用户在第二天有可能购买的商品。
 - 2. 作为团队(2人)主要参与者进行了特征分析、模型建立、算法实现等工作,最终取得第一赛季Top 10%的成绩(719/7186)