

戚伟杰 (男, 24, 汉族, 中共党员)

中山大学 · 计算机科学与技术 · 本科 · 现居深圳 · 一个月到岗
手机: 13580491531 · 邮箱: 568445851@qq.com · 172cm · 65kg · 未婚



工作经历 深圳百分点信息科技有限公司 | 算法工程师
2015-07 → 2017-02 1、负责利用利用大数据相关技术对业务进行数据分析、挖掘产品特征、用户行为特征，构建用户精准运营的指标体系；
2、参与大数据平台的研发，在 Hadoop、storm、spark 上进行算法开发。
求职方向 数据挖掘/算法研究实现/数据建模/系统架构

时间线

2016-12 → 2017-02 pangu 网页关键信息抽取

格式化网页抽取特征，训练神经网络模型去预测 dom 树的路径为网页关键信息的概率，从而实现网页关键信息的自动化抽取和模板的自动生成，fbeta≈90%。

- 调研网页抽取的现在方法，结合机器学习的热潮提出使用模型来代替人工指定规则的繁琐；思考设计特征和实现部分特征的特征提取；
从 es 获取网页数据，定义路径、base 表、模板等概念，人工标注模板和生成 base 表供生成样本；用 django 搭建整个 web 应用框架。

2016-09 → 2016-11 xnr (虚拟人) 项目模糊拉通算法开发实施

对多种数据源数据进行归一化处理，使之符合模糊拉通的数据输入条件；对各类 ID 进行建模分析，找到可以合理拉通的 ID 集合，配合精确拉通，把一个人的所有 ID 尽可能的准确拉通，提供一个虚拟人的全面视角。

- 独立开发，结合客户的数据情况，实现模糊拉通算法从 hive 源数据清洗、特征选取&抽取、训练模型、模型预测，预测结果汇聚的整套流程的工程化，实现了流程的可配置性和训练模型的自动化；难点在于算法和交付的结合，也就是工程化（可配置性和自动化）。

2016-03 → 2016-08 gemini 模糊拉通算法开发

从海量用户行为中抽取有用的数据 (ETL)、进而抽取有用的特征，用 GBDT 决策树模型对不同终端的 ID 之间预测可能是同一用户的 ID 属于同一个用户的概率，融入改进的最大联通子图算法，对模型预测的拉通结果去噪和汇聚，最终开发了模糊拉通算法。

- hive 数据 ETL，检查出公司数据采集的上的很多问题，推进了数据采集标准的指定，得到有用数据；分析得到对模糊拉通有用的特征，提高模型训练指标，推进项目进度；
设计和实现适用的最大联通子图算法，优化了模糊拉通的质量；

2015-07→2016-02 proxima 分布式实时算法平台

推荐算法很多都是离线计算，proxima 使用 kafka+jstorm+redis+zookeeper 实现了推荐系统的准实时性，kafka 接入数据，jstorm 提供网络 topology，zookeeper 负责调度，redis 负责存储结果；proxima 设计了一套算法 api，可接入各种流式算法，例如 CF 和 lr；后续开发个性化配置，可以针对不同的客户配置不同的算法和推荐策略，提高系统的实用性。

- 前三个月阅读 proxima 系统源码，熟悉组内产品和基本开发工具，同时实现一些小功能；
10 月份开始实现 proxima 的个性化配置部分，实现了 proxima 的动态可配置性，使得 proxima 的鲁棒性更强，配置更简单，使得动态可以调整整个推荐系统而不用重启系统。

2015-04→2015-06 百分点实习

支援 tcl 项目的 elasticsearch 后端查询接口开发
用户画像项目试运行维护，学习 hadoop 集群使用和熟悉公司产品

2011-9 至 2015-7 大学教育

中山大学信息科学与技术学院 计算机科学与技术系网络工程专业
本科 学士 | 班级名次：专业前 10 名

校内项目简略如下：

2014-11 → 2015-03 网络购物平台：实现一个类似于淘宝的购物平台

2014-07 → 2014-08 小面包搜索：对监控网站进行 key 搜索，返回包含 key 的 url

2014-06 → 2014-07 搜索引擎：实现一个小型的微博内容搜索引擎。

技术

实战项目

Python

Pangu、xnr、gemini、毕业设计、课程大作业（信息检索、小面包搜索）

CF 算法

Proxima

决策树算法

Gemini, xnr

Django

pangu, 毕业设计

Java, jstorm

Proxima 的个性化配置系统

Hive

Gemini, xnr

至少知道以下算法

数据挖掘算法

Knn, k-means, ID3, C4.5, CART, GBDT, RandomForest, SVD, SVM, apriori, LR, HMM, EM

兴趣特长

兴 趣
爱 好

健身 篮球 象棋 跑步
听歌 发呆 敲代码

所获证书

大学

大学英语四级
大学英语六级

校内职务

2013-9 至今

计科三班党支部党支书

2012-9 至 2013-7

文体委员

2012-9 至 2013-6

联想俱乐部部长

2012-4 至 2012-10

院党支部宣传委员

校内奖励

2013-04

优秀学生干部

2012-10

中山大学健康之星

2012-07

单项奖学金

2014-04

优秀团员

公益活动

2011-10-2

广东科学中心

为广东科学中心提供公益志愿者，引导小朋友在广科内快乐玩耍

2011-11-26

逢源街文化活动中心

“同一片蓝天”逢源街新广州人子弟社区共融联欢日公益活动

2012-4-29

广州五号停机坪购物广场

为首届广州白云时尚音乐节提供志愿者服务，引导人流

2014-4-20 到 25

国家超级计算机广州中心

后勤组网络维护志愿者，为 ASC14 竞赛提供网络服务

我对自己的第一评价是乐观，与周围的人相比，我觉得自己对事情的发展总是朝着积极乐观的方向思考，起码这能让自己总是处于比较好的状态去应付一切。第二印象是积极上进，不求最好，但求更好，在生活中的很多方面会保持主动的姿态，对外的表现可能是：一个闲不得的人第三印象可能是善于思考，善于总结反思，善于推理类比。

自我评价

觉得自己最大的不足是喜欢做真实的自己，不喜欢圆滑，喜欢直来直去，不懂得掩饰自己对某类事情的厌恶，可能不可能和自己讨厌的人呆在一起吃饭，总之，我喜欢的人喜欢我，我不喜欢的人也知道我不喜欢 ta，再次可能是有一点点的白日梦情结，这可能跟自己的乐观心态有关，事情都喜欢往好的方向想，但现实哪有这么好呢？我自己会很快的调整自己的目标，喜欢旅游或者读书，因为这样，心灵和身体至少有一个在路上。