

段义波

住址：北京海淀五道口

电话：86-18601730891

邮箱：dybwall1234@gmail.com

教育背景

广东外语外贸大学

统计学专业

2007-2011

英语：六级

生日：1988-10

个人技能

- 熟悉数据开发，有运营数据分析挖掘、数据平台开发，数据可视化，搜索推荐等经验。管理招聘经验。熟悉数据项目设计和交付。
- 熟悉 Spark, Hadoop, Hive, Linux, Storm, Kafka, Elasticsearch 等数据存储及处理程序。
- 熟练使用 Python, Java, R。熟悉 Scala, Shell, Sed, Awk。熟悉运维和分布式架构设计。
- 无障碍阅读英语文档。

工作经历

学堂在线

数据开发

2016 年 5 月—2017 年 7 月

公司简介：教育部和清华大学直属的在线慕课网站

职责：1，教学分析平台开发。与产品经理共同谈论需求，综合考量老师对学生学习数据的报表需求，组织成员进行开发。

2，搜索项目产品规划与开发。场景预调，包括搜索词提示，结果干预等。然后搜索系统的架构和开发。

3，军委搜索项目的开发，维护。

百分点信息科技有限公司

研发工程师

2015 年 1 月—2016 年 5 月

公司简介：大数据解决方案提供商

职责：1，公司 Spark 集群小组负责人，公司内部技术学院讲师，负责公司 Spark 技术分享和培训。

2，一键式大数据系统集成部署产品的团队管理，开发，实施。

3，对接文本挖掘平台到大数据集成平台，以及一些文本项目的开发实施。

京东金融(离职原因：跟随领导去了百分点) 数据挖掘

2014 年 8 月—2015 年 1 月

职责：1，根据用户订单及浏览等数据，对用户进行画像，整理已有画像字段，部署用户画像任务。

2，参与保理业务挖掘项目，使用 Spark 图计算进行用户分组。

3，白条推荐的效果收集及利用 R 自动生成邮件报告。

天天动听（阿里巴巴旗下公司）

数据工程师

2011 年 8 月—2014 年 8 月

职责：1，用户行为分析，分析用户听歌行为，并使用 R 进行可视化。

2，管理数据以及数据接口，为运营，产品，BD 各部门提供数据支持。

3，数据平台的搭建维护，使用 Hive 存储数据，处理用户日志，做数据汇总，主要负责 CDN 相关数据和搜索相关数据。

4，用 Scala 语言编写基于 Spark 的 K-Means 算法，基于收藏歌曲的用户对歌曲进行聚类。

项目经历

军职搜索项目开发

项目职务：数据开发

所在公司：学堂在线

项目简介：为军职慕课平台提供军职课程搜索 API

项目职责：需求分析，数据流架构设计。使用 Python 开发，从军职课程 Mysql 数据库拉取数据建索引到 ES，然后完成搜索，加权排序等功能。

教学分析平台课程数据触发式下载

项目职务：数据开发

所在公司：学堂在线

项目简介：使用 spark，Kafka 完成表格数据的触发式下载

项目职责：需求分析，数据流架构设计。使用 Python 开发，用户提交下载请求，Kafka 收到消息后，触发 spark 从 hive 中提取数据，进行数据聚合，生成 Excel，提供下载链接。

金融舆情分析

项目职务：数据开发

所在公司：百分点

项目简介：使用 Storm，NLP 等技术对新闻和股吧数据进行分析

项目职责：需求分析，数据流架构设计。分别用股价涨跌，人工打标，正则匹配等方式给新闻数据打标，训练出三个模型。配置为 Thrift 服务，用 Storm 方式调用情感分析 API 进行新闻数据的打标，然后把数据经过一些归一化处理，灌入 Mysql，作为回测数据。

相似邮件搜索

项目职务：数据开发

所在公司：百分点

项目简介：通过消息数据的搜索返回类似消息

项目职责：需求分析，架构设计。数据的预处理，插入 ES，词性过滤，通过 DF，W2V 等方式扩展权重，通过搜索 ES 返回相似消息。