

## 求职意向：搜索、推荐算法

- 姓名：严玉良
- 性别：男 出生年月：1990.4
- 手机：15267851139 邮箱：yyl8781697@live.com
- 个人博客：<http://kubicode.me/>

## 工作经历

**2015.10~今：阿里巴巴 岗位：算法工程师**

### 1. 内容搜索（2015.12 ~ now）

承接阿里集团内非宝贝相关的搜索，比如帖子、清单、视频、话题等，支持场景有淘攻略、微淘、头条、小蜜等，同时还参与双十一内容模块的活动（日常两千万级别的pv，活动期间五千万级别）。该项目中我主要负责排序算法（亿级别的文档）及其优化工作，包括常规排序特征、相关性、个性化、ctr预估等。同时还负责该项目的全部后端工程的开发和运维，比如算分插件的开发、搜索业务规则、日志的搜集统计、线上监控管理等。

### 2. 知识图谱-数据挖掘（2015.10~2015.12）

项目主要是为集团内建设底层的知识库，我主要完成同义词挖掘、属性归一化、新词发现等功能。

**2015.3~2015.10：IBM研发中心 岗位：java工程师（实习）**

### 1. BuildForge项目(2015.3~2015.10)

BuildForge 是一个基于web的构建和发布管理系统，用于在软件开发周期内帮助开发团队自动控制，追踪，分析复杂的构建和发布过程。我主要负责Agent端的执行调度及通知优化，Java引擎端的功能开发，支持不同的执行命令及与其他语言版本的引擎功能保持一致。

## 主要算法应用

### 1. 2017.8~2017.11：Wide&Deep个性化模型

应用于内容搜的个性化模型。user、query和doc三路各自进行dnn之后进行concat作为deep的输入(a)，同时还将doc的行为提取连续特征作为单个dnn的个输入(b)，最后还将用户行为和doc自身的内容进行FMLayer作为Wide层(c)，最终a、b、c的输入进行加和之后使用sigmoid激活。该模型使用TF进行训练，自己使用c++重写前向预测。双十一期间上线之后ctr提升了5%。

### 2. 2017.3~2017.7：DSSM语义相似度计算模型

应用于内容搜环境的Query-Title语义相似度计算。使用两个trick:1.根据文本相关性降低正负样例的噪声,2.基于分词gram(segwordhash)的Word Hashing。基于6亿的淘攻略用户搜索日志在PS上训练标准参数的DSSM（NDCG@10能到0.88），其语义分数比传统的Match方法具有非常大区分性，上线之后ctr提升了3%。

### 3. 2017.2~2017.3：基于MAB的多资源混排

用于淘攻略的混合排序。不同资源无法在单一链路里面进行排序，基于此根据用户在不同query下的行为制定混排策略，利用MAB模型，以点击率为优化目标，将每个坑位视为一个Arm,其策略就是各个资源类型 $R_i$ ，为了考虑exploit和explore，同时使用了Thompson sampling和Epsilon-Greedy两种方法。该方法灵活简单高效，最后线上ctr提升了3%。

### 4. 2016.10~2016.12：GBDT文本相关性

搜索环境下的文本和帖子内容相关性。利用外包的评测数据作为训练集，以bm25、tfidf、词性、语言模型为特征（均为bit存储）用GBDT进行训练,由于线上latency的需求，调参时折衷了算法性能和模型复杂度，最终版GBDT在Benchmark的NGCG@10提升了3%。

### 5. 2016.3~2016.5：Query丢词改写

对于非高频Query，根据核心词、pv量得到候选丢词，再根据用户行为构造训练pair对，用LR模型以CTR为目标得到top1丢词结果，该丢词算法提升了5%的召回率和1%的ctr。

#### 6. 2015.11~2015.12: 同义词挖掘

基于用户搜索点击行为进行同义词挖掘, word2vec剪枝, LR/GBDT训练, 评测准确率93.64%, 为集团增加了一倍的同义词.

#### 7. 2015.10~2015.11: 属性归一化

将近义或者不同写法的属性进行归一, 转为二部图上Attr的相似问题, 最后使用LR进行训练预测, 准确率90%.

## 技能

- 算法技能: 常用回归分类方法(LR、Svm、RF、GBDT等), 常用搜索技术 (LTR、、CTR预估、相关性等) 以及主流的深度学习算法.
- 常用语言: Python、Shell、C++、Java、Lua
- 常用框架: Hadoop (&Streaming)、TensorFlow、Spark
- 常用开发工具: Vim、Mac

## 教育经历

- 2013.09-2016.06 宁波大学 | 计算机应用技术| 研究生(推免)--研究方向: 数据挖掘
- 2009.09-2013.06 宁波大学 | 数字媒体技术| 本科生

## 论文/专利/著作权/奖励

- [“FSMBUS: 一种基于Spark的大规模频繁子图挖掘算法”](#),《计算机研究与发展》(EI)
- 专利: 商品搜索中一种资讯辅助购物决策的个性化导购技术,严玉良,\*\*\*,\*\*\*(国家在审)
- 软件著作权: 基于.Net的Ajax直接访问类库的框架中间件, 2014SR120294[\[github\]](#)
- 软件著作权: Goal Anchor理想氧吧网站管理系统, 2012SR016235
- [2014阿里巴巴大数据竞赛天猫推荐算法挑战赛决赛排名12/7274\(队长\)](#)
- 全国电子商务挑战赛特等奖、省多媒体大赛网站类一等奖 (2次)、二等奖、三等奖, 省创业计划挑战赛三等奖
- 校十佳学生, 宁波市优秀大学生(15人), 校十佳创业之星、国家奖学金、推免奖学金、校特等、一等奖学金