

赵明星

Golang, Python, C++

(+86) 188 1086 0130
zhaomingxingDL@gmail.com
igoingdown.github.io
igoingdown



研发工程师

教育经历

- 2016-2019 北京邮电大学, 网络技术研究院, 网络与交换技术国家重点实验室, 硕士研究生.
2012-2016 北京邮电大学, 计算机科学与技术学院, 本科.

获奖经历

- “编程之美” 2017 亚军, 2/1118
- 百度 & 西交大大数据竞赛优秀奖, 10/1393
- 全国研究生数学建模竞赛二等奖
- HackPKU 2017 车道线检测赛题冠军

工作经历

2019.09-今 字节跳动, GIP-热点, 服务端开发工程师.

热榜聚类系统负责人

设计基于面向热榜业务的聚类算法: 设计并实现基于关键词的聚类系统, 彻底解决基于 k-means 的聚类算法的漂移问题; 较基于 k-means 的聚类准确度提升 30%, pct99 降低 90%。

提升接口性能: 精简存储模型, 只存聚类 embedding 和少量 tag, abase IO 降低 60%。

解决聚类冲突: 实现冲突检测并对大 cluster 进行聚类豁免, pct99 降至 240+ms, 彻底消除了冲突。

热榜计算系统负责人

负责热榜 2.0 实时热度计算模块设计与实现, 完成并持续优化热度计算模块, 提升榜单内容质量。

提升榜单稳定性: 设计并实现 snapshot 机制, 榜单出问题支持锁榜, 降低用户感知; 针对文章计数不可回溯的问题, 设计并实现极速版热榜热度计算器, 作为线上服务的临时备份。

提升接口性能稳定性: 抽象并实现多级流式处理系统计算热度, 多流程全并发, 榜单更新频率从 10min/次降至 1min/次; 通过增加 cache 的方式, 降低 abase IO, 接口性能提升 90%

竞品 push 监控系统负责人

独立建成竞品热点监控系统, 覆盖竞品站点, 账号, 和 app, 提升头条热点内容全网覆盖率 25%。

独立建成可插拔全网竞品 push 监控系统, 监控全网热点主要竞品首发率, 头条热点 push 首发率提升 40%。

设计并实现通用的并行和分批处理 SDK 供内部使用, 线上代码仓库接入量 20+。

热点审核系统负责人

提升热点内容审核时效: 由 0 到 1 独立打通热点运营和内容审核, 建设热点审核的准实时专项数据流, 数据产出延迟平均小于 2 小时, 产出审出率、可热率和延迟并基于上述数据构建报警系统。

独立设计并实现热点和内容的相关性判定接口, 使用异步缓存支持接口 QPS 8000+。

2019.07- Shopee@Singapore, Platform Engineering Group, Software Engineer.

2019.08 设计并实现 RDS 管理平台。通过创建 RDS 时自动添加透明 Slave 节点降低故障恢复时间。

- 设计并实现 RDS 备份系统。将 task 生成器, 调度器和执行器分离, 提高系统可扩展性。投入使用后, 降低 DBA 团队 70+% 工作量。

实习经历

2019.01- 字节跳动, UGC 研发部, 后端研发实习生.

- 2019.06 将单个 ES 大索引按时间拆分为多个小索引, 保留近一年的数据。查询性能提升 30%, 磁盘用量降低 50%。
- 基于 kafka 和 redis 重构热点事件相关指标监控系统, 相比基于 crontab 的系统, 任务失败率降低 20%。

2018.05- 微软亚洲研究院, 研究实习生.

- 2018.07
- 在真实的 Excel 数据集上, 设计并实现 **SeqGAN** 回收表中有可能被用于生成 chart 的列的组合, 在 recall 为 82.4% 时数据集中正样本数量增加 65%。
 - 在 SeqGAN 中使用基于 **Monte Carlo 搜索** 的方法代替仅仅使用最终生成的组合与用户用于绘制 chart 使用的列的组合的相似度作为强化学习中的 Q-function, 提升了 action 的估值准确率。
 - 在真实的 Power BI 数据集上, 基于 TF-IDF 算法构造加权词向量特征, 设计并实现类 **VGG** 模型预测用户可能生成的 chart 的类型。precision 比加入 RNN 模型高 13%, 比最好的 ML 模型 (RF) 模型高出 8%。

项目经历

2017.05- “编程之美” 2017 挑战赛, 亚军.

- 2017.08
- 资格赛使用 pytorch 实现融入 **attention** 和 **copy 机制**的 seq2seq 模型, **MRR > 66%**, **排名 13/1118**。
 - 初赛使用 C# 语言, 基于 luis 和 **bot framework** 设计并实现能够回答北邮相关问题的问答机器人, **排名 3/200**。
 - 决赛阶段为初赛设计的问答机器人添加了“找工作”、“找对象”等实用功能, **决赛排名 2/8**。
 - bot 开发中, 我负责**全部后端开发**, 主要包括单轮和多轮对话的设计、对话状态转移、信息存储和检索等。使用 python 实现爬取北邮人论坛的**实时爬虫**, 基于 **TF-IDF 实现搜索引擎**。

科研经历

Dec. 2017 **参与实验室项目组隐私保护的关键研究**, 在项目组中负责实现图和序列上的**差分隐私算法**的实现, 在多个真实数据集上进行实验, 在隐私约束下数据可用性达到业界领先水平并与导师合作发表论文.

X. Cheng, S. Su, S. Xu, L. Xiong, K. Xiao and **M. Zhao**. A Two-Phase Algorithm for Differentially Private Frequent Subgraph Mining. **IEEE Transactions on Knowledge & Data Engineering**, 30(8):1411-1425, 2018