

卓俊杰

• 联系电话：(+86) 138-1626-7895

• E-mail : 13816267895@163.com

• 知乎：学习学习再学习



专业技能

- 熟悉自然语言处理，包括聊天机器人，搜索推荐、文本分类，NER 和知识图谱等；
- 熟悉多模态模型：包括 VedioBert, CLIP4CLIP, CLIP, VLM 等模型
- 熟悉深度学习框架及开发模块：Keras、Tensorflow、Numpy 等；
- 熟悉 NLP 及机器学习基本算法：BERT、Transformer、LightGBM 等；

个人评价

- 本人爱好读书，热爱 NLP。逻辑思维强，学习能力强，勇于面对挑战；对人真诚，对事负责；具有很强的团队协作能力和沟通能力，富有独立分析和解决问题的能力。

项目经历

2021.7-2021.9

基于用户历史行为的搜索优化

为提高客户在公司产品上的搜索 mAP，根据用户的历史行为对当前搜索结果进行优化重排

- 策略方法：分别采用统计法和模型两种方式，以适应不同阶段的用户点击数量，统计法适用于前期数据量较少的场景，模型适用于后期数据量多的场景。
 - 统计法：统计用户历史点击过内容的 metadata 喜好，对当前搜索结果进行重排；
 - 模型：采用双塔结构建模，左边是用户画像，右边是待重排的内容，推理时速度更快；
- 数据处理：通过 mysql 拉取用户埋点数据，构造（历史行为，搜索词，埋点行为）的正样本；负样本采用库中随机选取和暴露但是未点击的样本，达成模型整体不出大错又关注小细节的能力
- 效果：mAP 离线测试有 6% 的提升，上线有 5% 的提升。

2021.5-2021.7

腾讯广告算法大赛（多模态）模型

在腾讯广告算法大赛的多模态视频广告标签赛道中得分 0.81，排名第 18。

- 数据：共 5000 个广告视频，包括视频、文字、图片等特征，预测 81 类多标签分类。
- 建模：最终采用 Vedio Transformer 的整体框架，将四个模态通过[SEP]拼接起来，结合上 R-drop、MLP attention 等；中间分别尝试 Co-attention、Realformer 等。
- 落地场景：应用于内部视频多模态智能标签打标，方便下游搜索推荐等任务。

2021.8-至今

基于知识图谱的案例推荐

为客户推荐创意方做的案例，设计了两个实验（RippleNet、EGES+GraphSage），最终采用 EGES+GraphSage 的模型思维。

- 借鉴的思想是：GraphSage 可以用局部信息逐层汇聚来表示当前节点信息；EGES 可以将用户行为构建为同质图谱，以随机游走来得到正负训练数据
- 离线验证有 57% 的 mAP，待上线验证效果



2020.3-2020.12

多轮问答机器人

聊天机器人

为满足公司智能业务办公及面向客户的需求，开发一个智能聊天机器人，个人负责整个项目；融合检索、意图识别、多轮对话、知识图谱 KBQA 等多种方式搭建，并结合转人工服务。

- 思路解析：
 1. 检索和意图：通过粗排(Bool 检索和 BM25)、精排(Similar)的方式检索，返回最相似 QA 对；该 QA 对可能包含有意图(查账单、天气等)，结合对话管理 DM，构建意图决策树，进行槽填充式的多轮对话；KBQA 取用 1 百万个常识三元组；
 2. 推荐可能想问：当 top1 的相似度阈值介于中间，返回可能想问的问题选项；
 3. 闲聊模式：没匹配到任务型 QA，进入闲聊模式，用知识图谱 KBQA 和闲聊库检索。
- 效果：每天有 1w 个业务人员使用，有效问题达 75%，解决率 80%，知识库检索成功率 97%；推广面向 C 端客户。

2020.5-2020.10

金融风控知识图谱

知识图谱

为提高行内风控能力，结合 NLP 和知识图谱，用于贷前贷后的团伙欺诈、申请反欺诈等场景。

- 1、数据处理：数据来源为行内 2000 万客户基本信息，通过 sql 进行调用处理缺失值等；数据处理主要为相似度匹配：包括地址相似度匹配和地址有效性，公司名相似度匹配。
- 2、建模：采用 neo4j 分布式版本进行构建，最后节点总数为 1.8 亿，关系总数为 2.1 亿
- 3、应用：
 - (1) 团伙欺诈：根据 Louvian 算法（对比 LPA）进行社区挖掘，挖掘团伙，通过计算团伙中的占黑率，发现欺诈团伙，利用 d3.js 进行可视化；
 - (2) 贷后欺诈识别：根据业务经验指导，从图中发掘现有客户可能存在的欺诈行为：比如半年内进使用同一公司电话进件超 5 次，且关联的人中欺诈率超 60%。

2020.7-2020.10

贷后催收预测

金融风控

负责搭建信用卡贷后模型框架，包括 M0，M0+和核销模型。

- 项目业绩：1) 通过征信、总行等途径电话补缺，得到 1 万个名单，共回收 100 万。2) 核销阶段名单目前正在做 AB test。

教育背景

2016.9-2019.3

同济大学

机械设计及理论(深度学习)

硕士

- 核心论文 1 篇，专利 1 项

2012.9-2016.7

上海海事大学

机械自动化

本科

- 校级奖学金 2 次，优秀团员

社会经历

2021.5-至今

特赞信科技有限公司

算法工程师

2020.3-2021.5

浦发银行信用卡中心

算法工程师

2019.5-2020.3

上汽通用汽车

算法工程师