**个人简历**

**个人信息**

姓名： 陆鑫益 性别： 男

出生日期： 1994年11月1日 工作年限： 3年

手机： 18606224388 邮箱： 645096158@qq.com

**求职意向：算法工程师、数据分析**

**专业技能**

语言及框架：熟悉Python、SQL、Pytorch、Spark，掌握C++、Go

领域知识：搜索意图识别、NLP、数据分析、图像生成

**工作经历**

**2021.11-至今 腾讯微信事业群WXG搜索应用部 算法工程师 T8**

1. 负责垂类box意图识别及效果优化

微信搜索垂搜box对标百度阿拉丁，目前共140+类box，占大盘流量20%+，意图识别作为垂搜box准入的条件对线上用户体验以及下游流量分配具有重要的影响。

a) 搭建和维护大规模特征数据流 Spark开发

利用海量微信基础数据构建数十亿级别的数据宽表，包括query的曝光点击数据、query切换数据、doc基础信息数据等，每天稳定例行更新，支撑组内包括数据分析、意图识别、相关性计算、排序等团队对基础数据的需求。

b) 构建深度意图模型触发框架 Bert预训练、下游任务finetune

主导构建深度意图模型触发框架，包括微信垂搜场景基座模型预训练、意图样本自动打标、模型特征生成、模型结构优化等。针对垂搜特定场景，在微信搜索语料的基础上，加入垂搜图谱在预训练阶段注入更多的领域知识；基于大规模特征数据流，构建不同基础算子实现意图finetune阶段的样本自动打标；针对query短文本语义信息缺乏的问题，利用曝光点击doc实现语义增强；针对类目多且复杂的问题，基于预训练模型采用MMoE结构实现行业下多类目统一触发，大幅降低上线成本。

c) 探索使用语义大模型来解决意图问题 LLM大模型实践

基于现有的开源基座大模型（LLaMA、GLM等），使用微信搜索场景语料构造数据来对基座模型进行SFT指令学习，将原来独立的Bert意图分类和实体抽取任务转化为统一的开放场景生成任务，目前已经在大部分垂搜box上已经取得超越BERT的效果。

2. 探索表情图片生成

表情搜索是微信生态的特色内容，希望通过表情生成来扩充生态内的资源供给，同时探索文生图、图生图等能力的不同应用场景。

a) 基于stable diffusion，采用PEFT方式如Lora来快速训练不同主题、风格、人物形象的lora模型，通过prompt调用和融合不同的lora模型来扩充当前的表情数据库。

b) 结合ControlNet、OpenPose等模型，探索不同的应用场景，包括人物表情生成、动作姿态生成、一键换装以及特定风格迁移等，为新的内置功能做技术积累。

**2020.7-2021.11 腾讯PCG电商搜索算法团队 校招入职**

1. 负责商品搜索特征体系、精排模型的搭建和迭代 搜索排序

从零到一搭建搜索特征体系以及迭代精排模型，模型迭代路径包括DeepFM、DIN、TransFormer、ESMM等，线上CTR以及CVR均取得明显收益。

2. 参与QU模型的优化，包括搜索意图识别、Term实体识别、类目预测等模块

QU模型采用Bert结构，通过多任务的方式集成搜索意图识别和Term实体识别两个关键任务，有效提升各个任务的精度，其中Bert模型产出的Embedding还可以用于商品的向量召回。

3. 负责搜索数据的分析下钻、相关指标看板的建设和维护

统计分析用户行为数据，比如高频q-d对、用户点击偏好等来丰富商品、人群的画像内容，辅助模型个性化效果的提升。

**教育经历**

2017.9-2020.7

华南理工大学 控制工程专业 硕士

2013.9-2017.7

华南理工大学 自动化专业 / 金融专业 双学位学士

**自我评价**

3年搜索相关技术经验，对新技术抱有足够的好奇心，积极拥抱大语言模型。

技术栈较广，熟悉自然语言处理，了解图像相关知识，同时对于基本的大规模数据处理和分析具备一定的经验。

业余时间喜欢尝试量化模型以及读各类书籍。