马晓晨

手机： (+86)188 -1301-8781　　邮箱： maxiaochen@bjtu.edu.cn

教育经历：北京交通大学信息科学研究所（硕士）　　**辽宁工程技术大学 通信工程（本科 保研）**

Github Pages：laoma023012.github.io　　**Github：https://github.com/laoma023012**

专业技能：

1　基础开发能力：熟悉Linux环境，对于 Python **面向对象编程/多进程编程**（事业部开源多机多卡分布式语义匹配框架支撑） 具有工业级实践，具有海量数据处理经验，掌握基础大数据框架Spark，Hive，pySpark等（推荐项目基于pySpark，搜索项目基于Spark），具有**Scala / Java / Scala ->Java 混合编程（JDBC）实践经验**，对于**C++ / OpenMPI / NCCL / TVM / OpenCV** 等具有涉猎

2　基础算法能力：熟悉搜索推荐架构流程，对于工业界算法迭代具有明确的思路 搜索 1）搜索召回相关性截断&相关性排序：模型： DSSM -> Multi-View DSSM -> Term Weight Multi Loss DSSM 数据： 头部词基于行为采样，长尾词基于 **基础 NLP 信号** 采样 机制：基于探测机制的 Query 模型动态选择，基于精排萃取的 **Serving** 通用长尾模型，探测机制 2）搜索/推荐排序：埋点设计/日志清洗与前后端关联，**Feature**上线后端打点 / 线上线下AUC对齐，特征工程（用户/商品 交叉），多任务学习 DSSM+LR -> MOE 3）推荐显式/隐式召回：画像Tag显式召回，基于兴趣探测的元路径/随机路径游走，Node2Vector，Metapath2Vector

3　较强的学习能力和工作能力(本硕期间连续 **7** 年奖学金，事业部**季度个人奖，事业部最高绩效 S / S**)，扎实的**数理基础**(国家级数学竞赛，数学建模二等奖)，和**英语**能力(国家级英语竞赛二等奖，CET 6)，文献阅读能力和**口语**能力(BEC-V)

工作经历：

2019.7-2020.2 美团 垂直类App 搜索推荐组 常买清单 分类页爆品排序

一 基础数据管道搭

多粒度埋点设计**[Impression / Click / Cart]**，后端商品列表与前端行为关联，线上线下 **AUC** 校准对齐

二 基础排序模型上线

XGBOOST + LR，特征工程，用户特征（基于**Item2Vector，基础统计特征**），商品特征（基于统计），交叉特征（复购时间戳）等，基础正负样本构建，由于BASE是策略，上线初期 单展位 **点击率/访购率** 相对涨幅达到 **80%**，绝对涨幅达到 **20%**，后面由于组织架构调整到搜索侧

2020.2-2021.2 美团 垂直类App 搜索推荐组 搜索基础相关性(BS) & 排序算法

[项目获事业部个人奖 / 搜索满意度达到 96.6 / 配合排序侧 UV访购率至 2020 Q3 提升达到 绝对10%]

一 搜索基础相关性召回侧截断

1) 基础架构-线上，**分层搜索相关性架构**，高频词层级，长尾词层级，店铺/供应商/品牌 层级

1.1) 基于 **显式行为关系** 的 高频词 **DSSM** 模型（2020-Q1 / Q2）

高频词 基于显式行为 做 **真**正样本/**虚假**正样本/**真**负样本/**虚假**负样本 四个粒度构建，模型层面从 **biGRU -> Transformer** ， 词向量借助腾讯开源 **embedding** 向量，虚假正样本借助核心词扩展

1.2) 基于 **基础 NLP 信号** 的 长尾词 多视角 **DSSM** 模型（2020 Q3）

长尾词由于显式行为缺失，借助基础 **NLP** 信号 [Query 核心词 / 商品TAG / TAG 意图识别] 做采样，通过构建 **Query -> Tag 意图识别 -> Tag -> CSU 与 Tag映射 -> CSU 异构图 | 行为采样补充** 做二跳正样本采样，负样本初版完全随机负采样，后期根据Top词 **Query -> Tag意图识别 -> Tag，** 建立 Q -> Tag 真负样本池，模型层面在 **Query -> Doc** 层面采用多视角方案，引入基于 Term Weight 的 Q -> QC **[query核心词]** ，D -> DC **[商品Tag]**，Q -> DC，QC -> D，QC -> DC 共享参数多 **Loss** 方案

1.3) 基于 **泛相关性** 的店铺/供应商/品牌 策略（2020 Q1）

基于 字面意思 做泛相关性的 店铺/ 供应商/ 品牌的召回侧截断方案，避免误截断

1.4) 基于 **精排萃取** 的兜底 **Serving** 模型 (2020 - Q4)

基于相关性分数，综合排序分数，杰拉德相似度，拼音杰拉德相似度，意图识别，借助后期的人工标注 Label 训练精排模型，基于 *logits* 蒸馏 DSSM 模型

1.5) 基于 **探测机制** 的 **Query** 模型动态选择 (2020 - Q4)

1.5.1) 建立 Query级 A / B 显式效果监控

1.5.2) 基于探测机制的 Query 模型 动态选择，基于分层 显式监控 做 Query 和 模型 的动态选择，效果不佳的 Query 自动退化成 Old 模型

2) 基础架构-算法，工业级 **Job**分布式深度语义匹配训练解决方案（基于 **Horovod** 事业部内开源）

2.1) 面向 **Horovod OpenMPI** 的 多机多卡 数据分片（2020-Q1）

2.2) 基于 **异步编程** 的数据分发 消息队列 与 模型训练（2020-Q3）

2.3) 基于 **反射机制** 的模型Graph 与 数据分发 动态加载，Session Graph 分离（2020-Q2）

2.4) 基于 **Tensorflow** 移动端部署 Toolkit 线上部署裁剪方案（2020-Q2）

3) 基础架构-数据，搜索基础日志 **SearchView LabelMatch**

3.1) 基于前后端日志一致性的 Request-ID 关联（2020-Q1）

3.2) CSU -> SPU 日志建模与多粒度聚合 / 埋点修复等（2020-Q4）

二 搜索基础相关性排序分数透传&多目标学习 [RANK**组合作]**

2.1) 联合排序和相关性的多任务学习 **Shared Bottom -> Two Expert MOE**（2020-Q3 / Q4）

2.2) 多任务学习拆分成 **L0 相关性排序** 和 **L1个性化排序**（2020 Q3 / Q4）

三 搜索相关推荐隐式召回

3.1) 基于 **ANN LSH** 的 FAISS 索引召回 (2021-Q1)，访购率涨幅相对超过 20%

2021.2-2021.5 美团 搜索推荐组 基于图方法的兴趣探测的推荐召回 [补充召回，品宽提升相对 1%]

一 基于显式用户画像的首页Feed 推荐召回

1) 基础架构-图存储 Nebula

1.1) 实体：用户，CSU，品牌，类目，食材标签 **Tag**

1.2) 关系：用户-> CSU 偏好，用户 -> 品牌偏好，用户 -> 食材标签等，图存储 ***N*** 跳 召回

二 基于 隐式用户画像的异构图推荐召回

2) 基于隐式用户画像的 异构图 推荐召回

2.1) 基于 **Node2Vector** 的画像显式偏好无向二部图

2.2) 基于 异构 **User -> Tag -> 类目 -> CSU** 的 **MetaPath** 的 Graph Embedding

曾获奖励：

1. **校内&工作奖励**：美团事业部季度个人奖，最高绩效 **S / S**，连续三年北京交通大学校一等奖学金，连续四年辽宁工程技术大学奖学金，省政府奖学金，校毕业生典型人物（校Top15）
2. **实习经历**：快手MMU 多模态反色情算法，CETC研究院 NLP 算法研究，沪江网 CCTALK 智能客服
3. **基础学科竞赛**：全国大学生英语竞赛（研究生组）**国家三等奖，**全国大学生英语竞赛 本科生组 **国家二等奖,** 全国大学生数学竞赛 **国家二等奖** 美国大学生数学建模 **国家二等奖 等**
4. **专业学科&论文**：**IJCAI-2018** 阿里妈妈搜索广告转化预测复赛第二赛季: **Top1%，**航天星图高分软件竞赛 SAR图像地物目标分类 **Top1，**第十九届中国图形图像学会**最佳学生论文，发明专利一项 等**