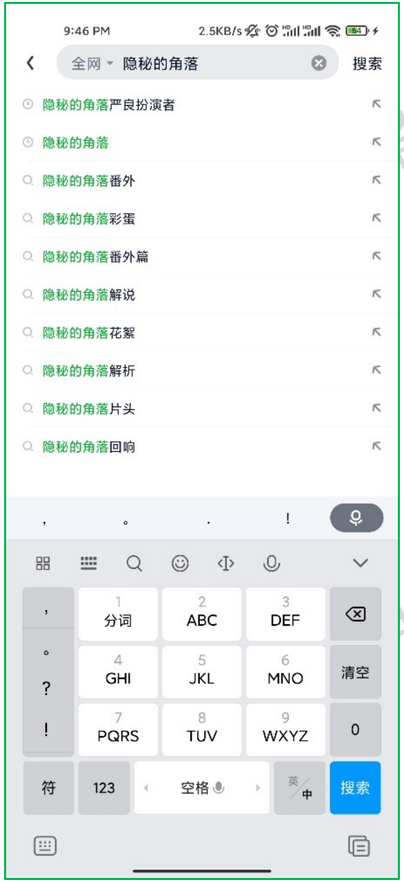
爱奇艺搜索排序算法实践

张志钢

iQIYI 爱奇艺
悦 享 品 质

搜索场景



多产品线

爱奇艺、随刻、极速版、TV等

多业务形态

综合搜索、各垂类业务

多数据类型

专辑、短视频、爱奇艺号等

业务目标

搜索效率

- Session CTR
- UCTR
- 二次搜索率

用户消费

- 用户播放时长
- 用户点击次数
- 用户互动次数

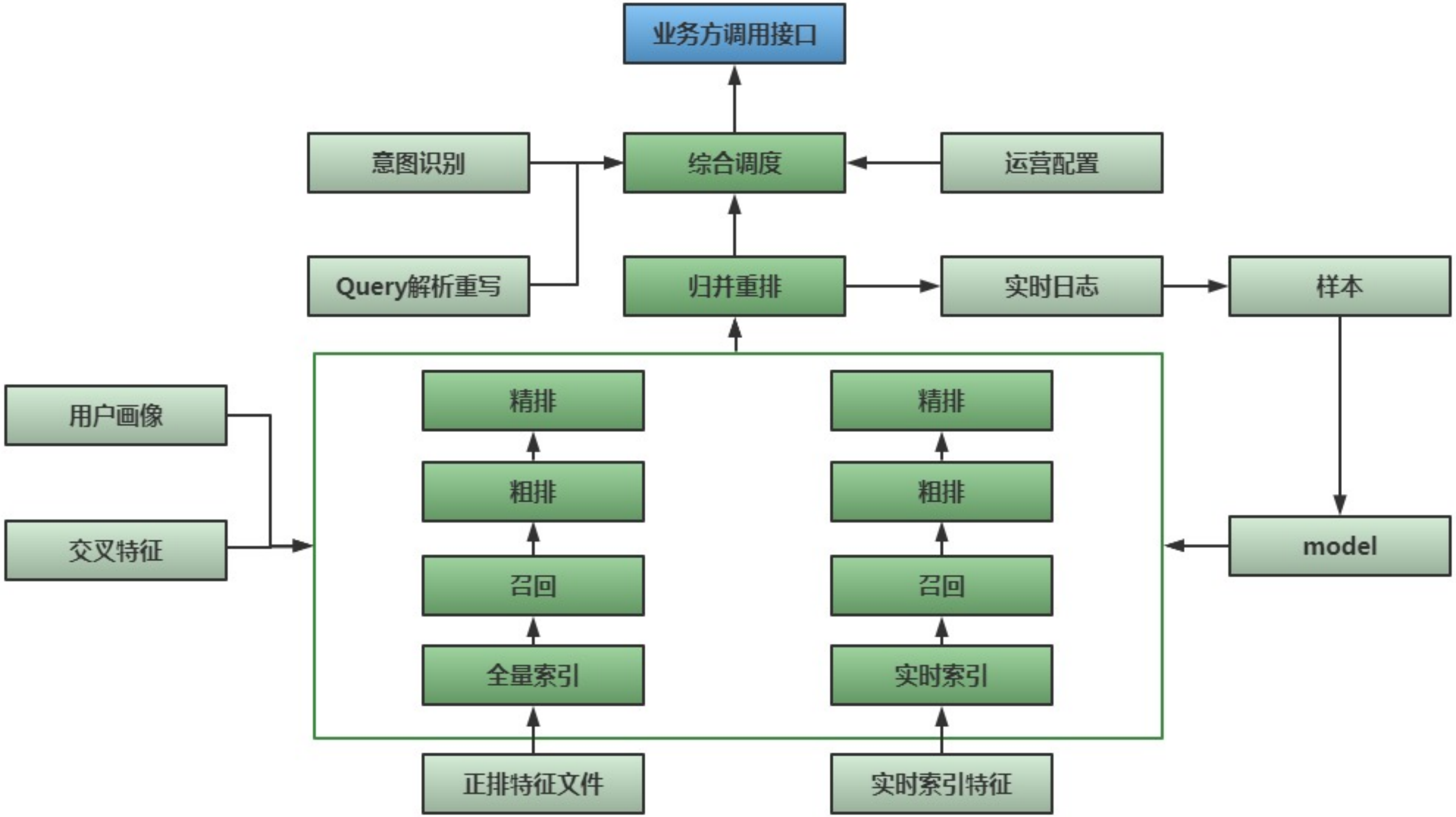
内容生态

- 全网视频
- 优质垂类
- 爱奇艺号

新热多样

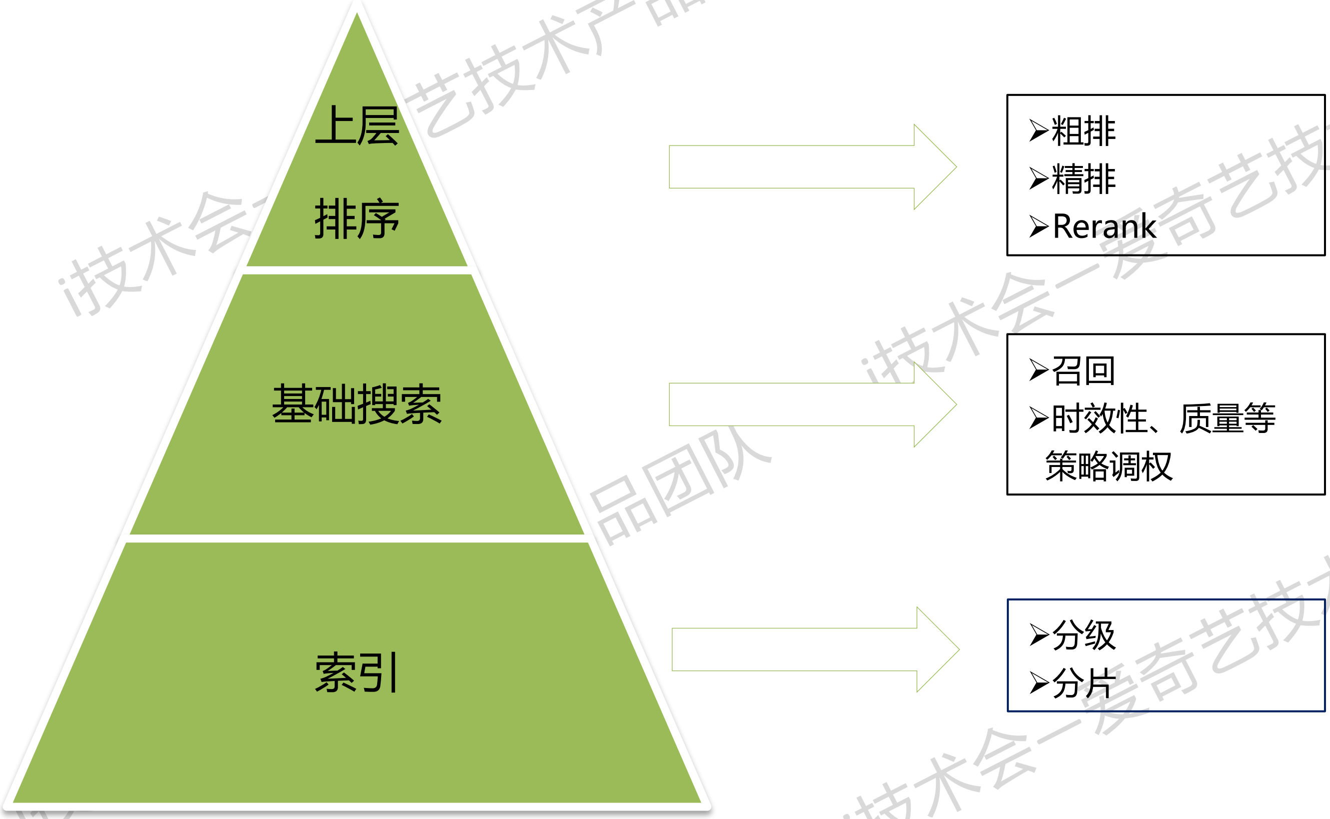
- 冷启动
- 时效性
- 多样性

整体架构



业务方调用接口
意图识别
综合调度
运营配置
Query解析重写
归并重排
实时日志
样本
用户画像
交叉特征
精排
粗排
召回
全量索引
精排
粗排
召回
实时索引
model
正排特征文件
实时索引特征

算法策略框架



排序流程


Corpus
Candidate Generation
Coarse-Grained Ranking
Fine-Grained Ranking
Rerank
亿级
十万级
千级
百级
十级

提要

召回
- 倒排索引
- Bert召回

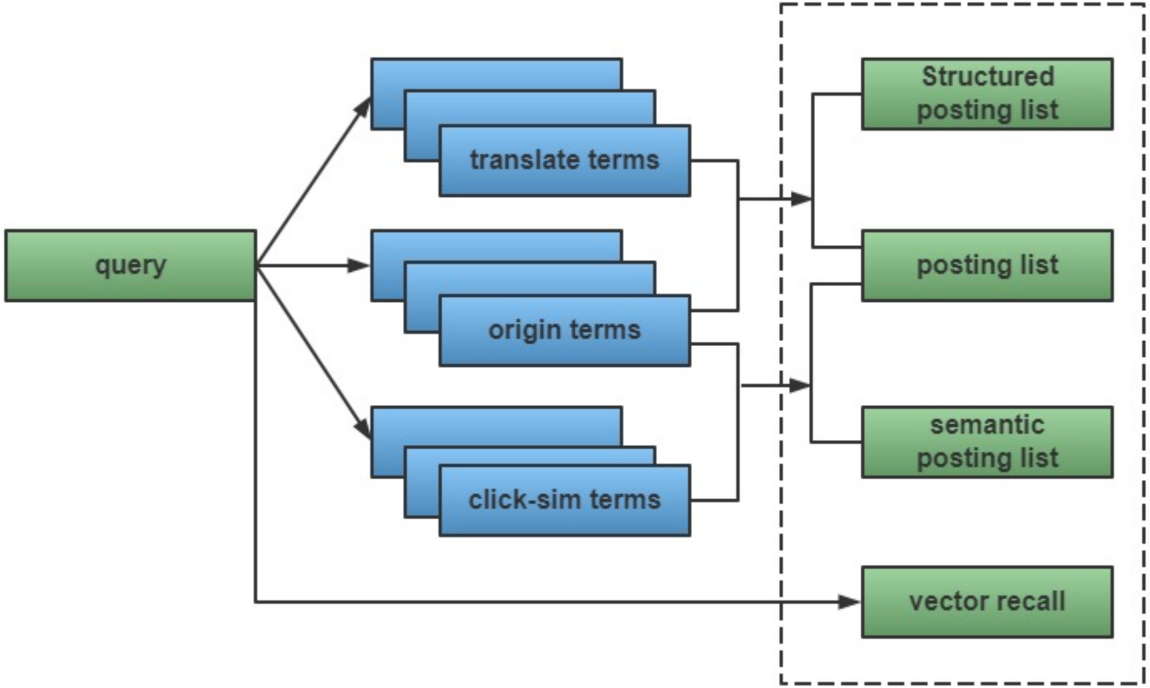
粗排
- 负采样
- 树模型

精排
- LambdaMart
- DNN升级
- TopK优化
- 多目标
- DeBias

重排
- SE-Rank
- PRM

可解释性
- 稠密特征
- 稀疏特征

召回阶段



平行语料

Query-Title点击对

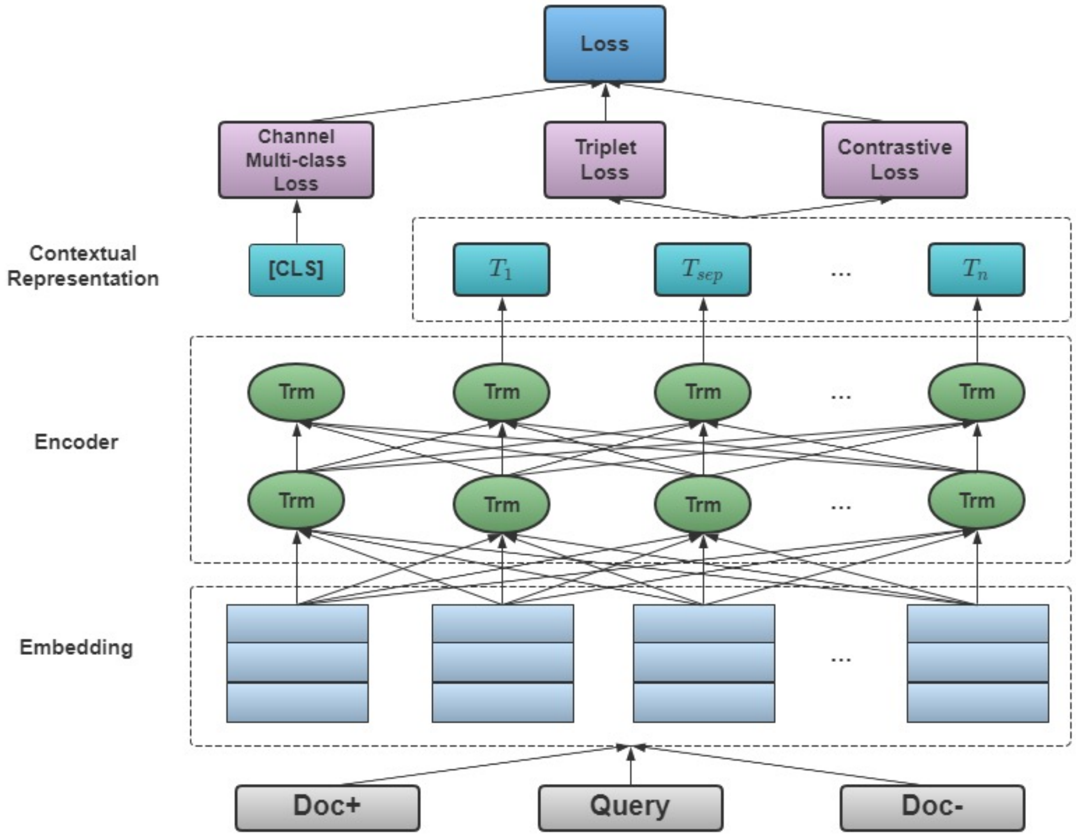
翻译模型

词对齐	短语对齐

排序筛选

特征抽取	模型排序

Bert召回



- 样本构建
- 输入特征
- 训练Loss
  - Triplet Loss
  - Contrastive Loss
  - Channel Multi-class Loss
- 在线服务

提要

召回
- 倒排索引
- Bert召回

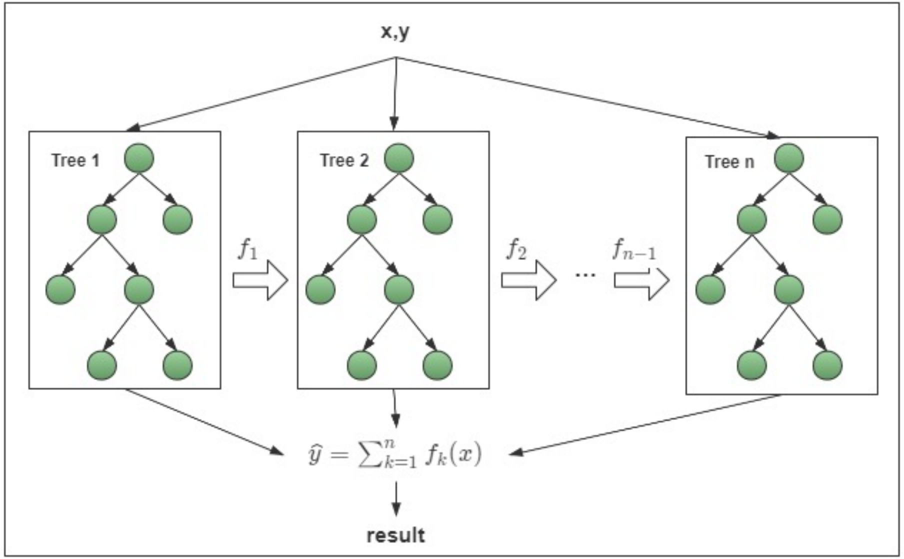
粗排
- **负采样**
- **树模型**

精排
- LambdaMart
- DNN升级
- TopK优化
- 多目标
- DeBias

重排
- SE-Rank
- PRM

可解释性
- 稠密特征
- 稀疏特征

粗排阶段



- 样本构建
  - 正样本
    - 用户点击
  - 负样本

- 重要特征
  - Doc质量特征
  - 交叉特征
  - 频道提权特征
  - 时效性特征

- 模型
  - XGBoost
  - 加权交叉熵

提要

召回
- 倒排索引
- Bert召回

粗排
- 负采样
- 树模型

精排
- LambdaMart
- DNN升级
- TopK优化
- 多目标
- DeBias

重排
- SE-Rank
- PRM

可解释性
- 稠密特征
- 稀疏特征

Learning to Rank

➢Point-wise

$$L_{pointwise} = \frac{1}{N} \sum_{i=1}^{N} (f_\theta(d_i) - y(d_i))^2 \qquad L = -\frac{1}{m} \sum_{i=1}^{m} y_i * log(\hat{y}_i)$$

➢Pair-wise

$$L_{pairwise} = \sum_{y(d_i)>y(d_j)} max(0, 1 - (f_\theta(d_i) - f_\theta(d_j)))$$

➢List-wise

$$L_{listwise} = \sum_{y(d_i)>y(d_j)} log(1 + (e^{f_\theta(d_i) - f_\theta(d_j)})) * |\triangle NDCG|$$

RankNet & LambdaMart

➢RankNet

$$S = f(x; w)$$

$$S_i = f(x_i; w), S_j = f(x_j; w)$$

$$p_{ij} = \frac{e^{(\sigma(s_i - s_j))}}{1+e^{(\sigma(s_i - s_j))}} = \frac{1}{1+e^{(-\sigma(s_i - s_j))}}$$

$$L_{ij} = -\overline{P}_{ij} log P_{ij} - (1 - \overline{P}_{ij}) log(1 - P_{ij})$$

$$P_{ij} = \frac{1}{2}(1 + s_{ij}) \quad s_{ij} = +1, 0, -1$$

$$\boxed{\lambda_{ij} = \sigma[\frac{1}{2}(1 - s_{ij}) - \frac{1}{1+e^{\sigma(s_i - s_j)}}] \quad s_{ij} = 1}$$

➢LambdaMart

$$\boxed{\lambda_{ij} = -\frac{1}{1+e^{\sigma(s_i - s_j)}} |\triangle Z_{ij}|}$$

➢LambdaRank重新定义梯度

➢训练框架采用MART

➢Label构造

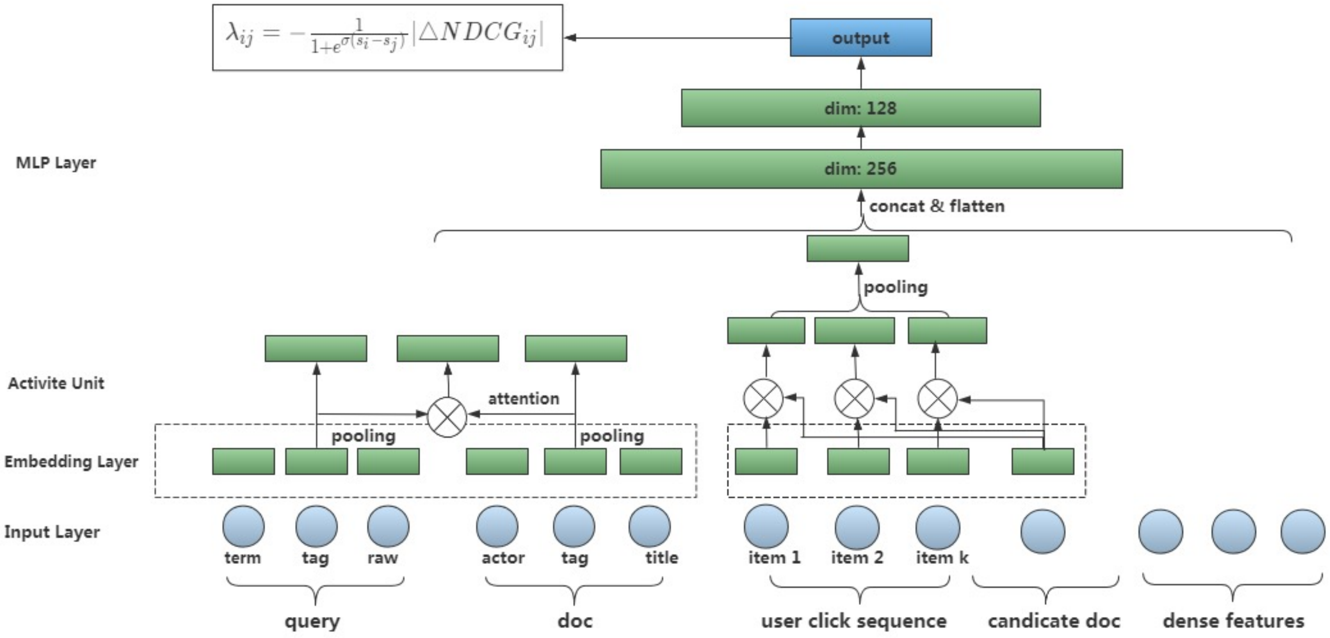
- ➢采用点击质量划分0~3共4档

➢适用于排序场景

- ➢对比大小而非绝对值
- ➢对于正负样本比例不敏感

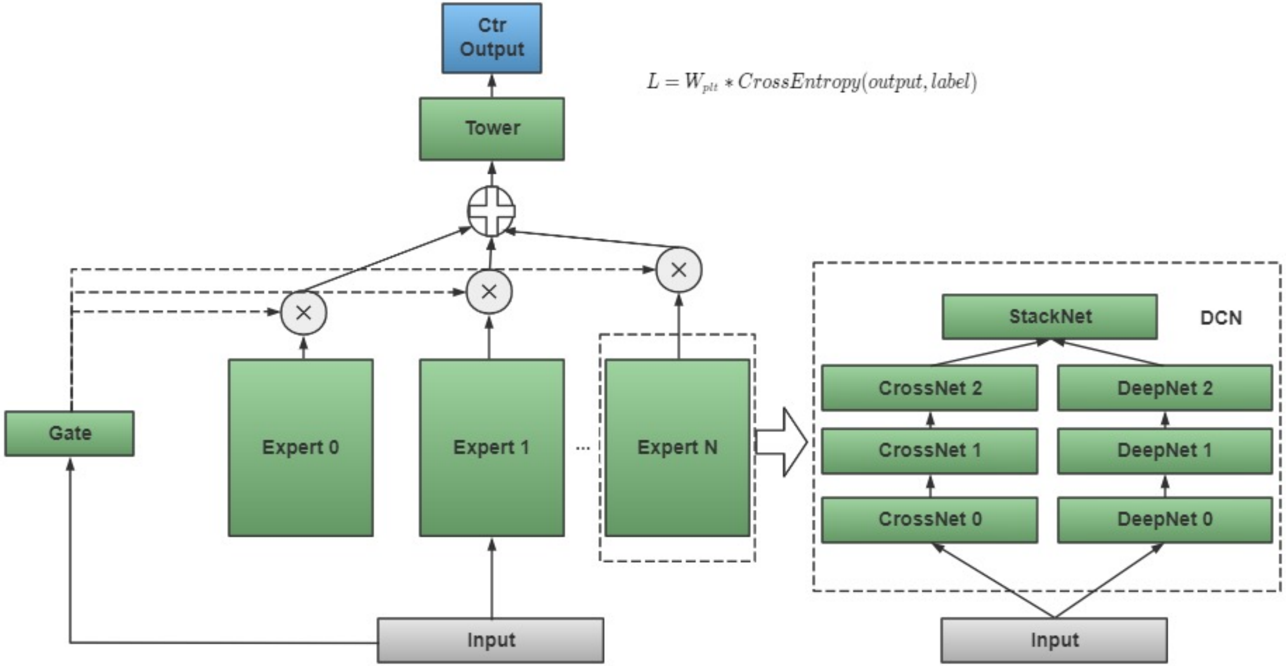
悦 享 品 质

LambdaRank DNN



- ➢样本采样
- ➢训练方式为List-wise
- ➢新增稀疏特征
  - ➢Query
  - ➢Doc
  - ➢User Profile
- ➢Loss
  - ➢LambdaRank
- ➢离线评估
  - ➢NDCG@5

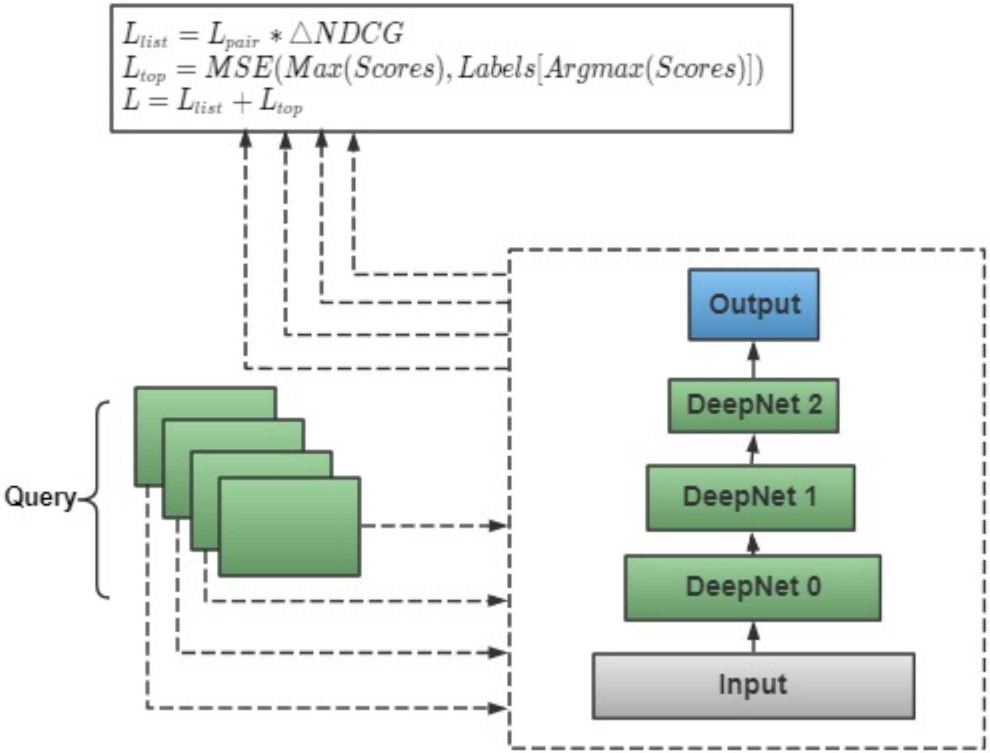
TopK Optimization – Hot Query



$$L = W_{plt} * CrossEntropy(output, label)$$

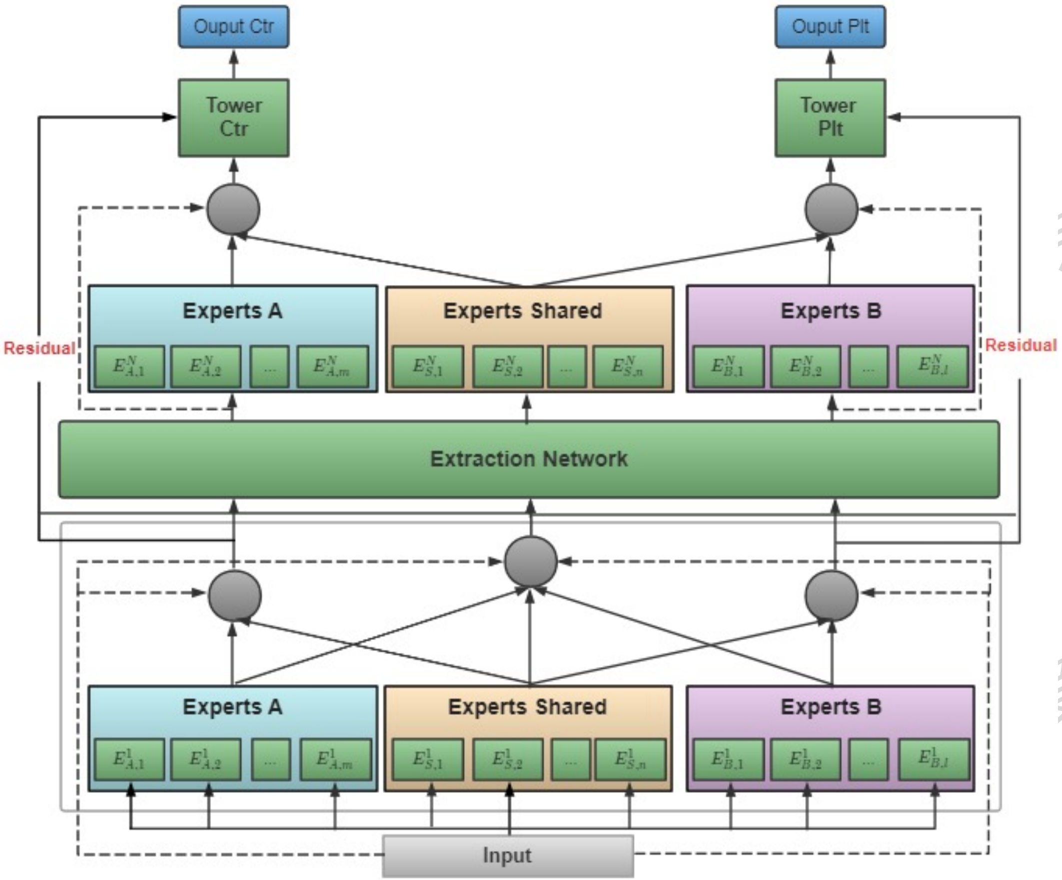
- 扩充大量头部样本
- 单目标加权
- 训练方式Pointwise
- 采用混合专家网络
- 新增ID特征

TopK Optimization – Longtail Query



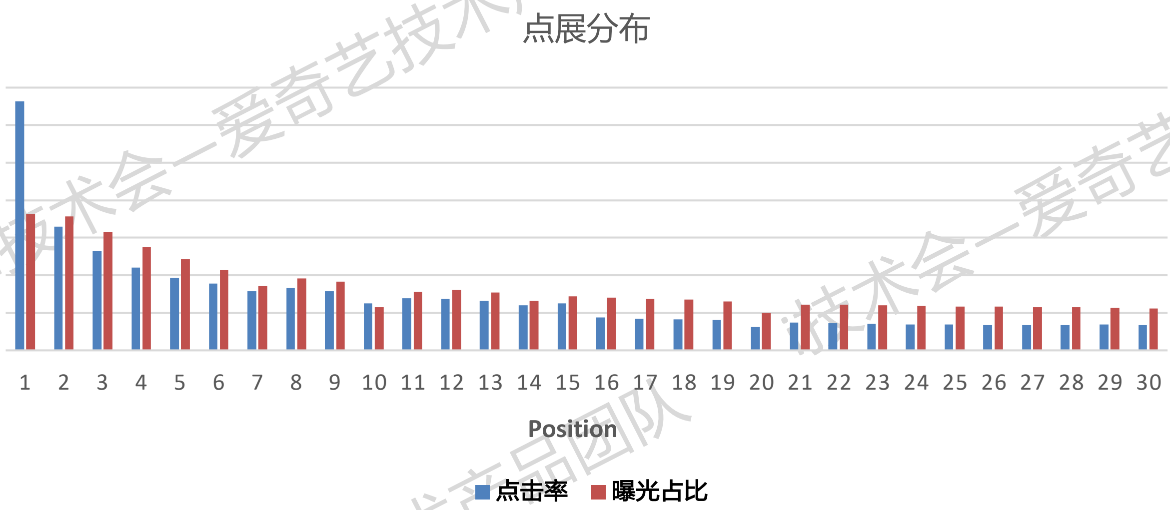
- ➢采用Listwise方式训练
- ➢限制单条样本下Doc数最小为2
- ➢Label采用点击质量
- ➢加大对Top1排序以及Label绝对值偏差的惩罚

Residual PLE



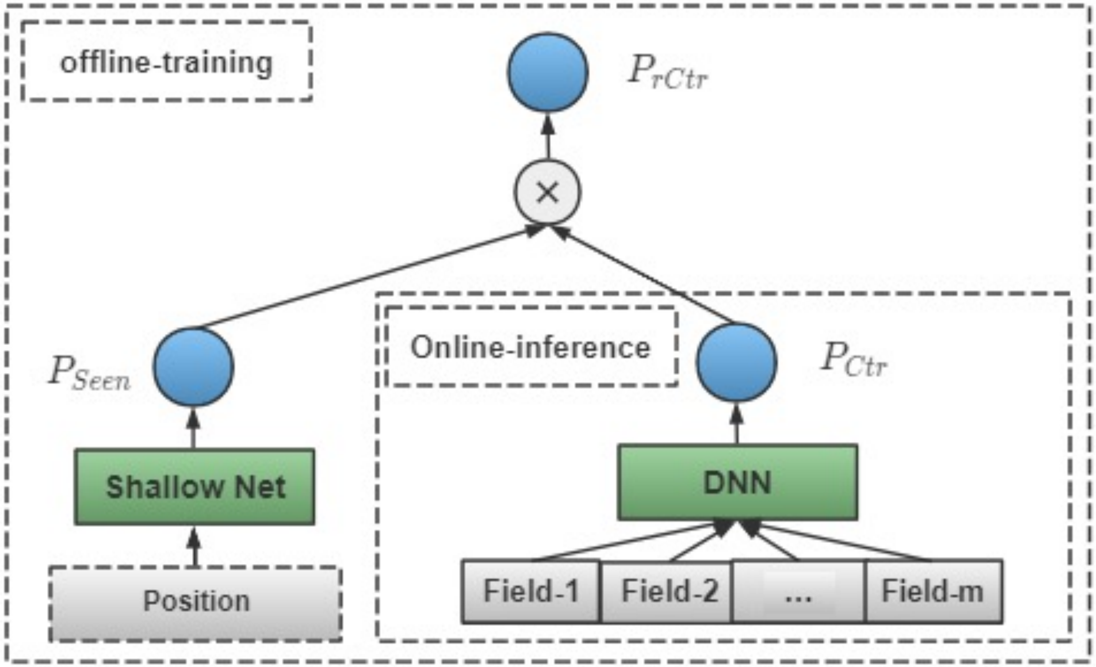
- 时长进行Log平滑处理
- Loss采用
  CE(Ctr) + alpha * MSE(Plt)
- 单个Expert采用DCN结构
- 双层PLE，任务塔加入首层残差
- 线上融合：
  $\mathrm{Norm}(P_{Ctr})$ * ($\mathrm{Norm}(P_{Plt})$ + Bias)

Position Bias



- ➢Position及少量特征，构建单独塔，输出加入到模型Wide部分
- ➢Drop-out 0.15
- ➢线上固定Position特征值为1
- ➢AB指标负向

Position Bias



- Position及少量特征，构建单独塔，建模被用户看到概率P_{Seen}
- 真实点击率预估值
  $P_{rCtr} = P_{Seen} * P_{Ctr}$
- 线上预估只取P_{Ctr}

提要

召回
- 倒排索引
- Bert召回

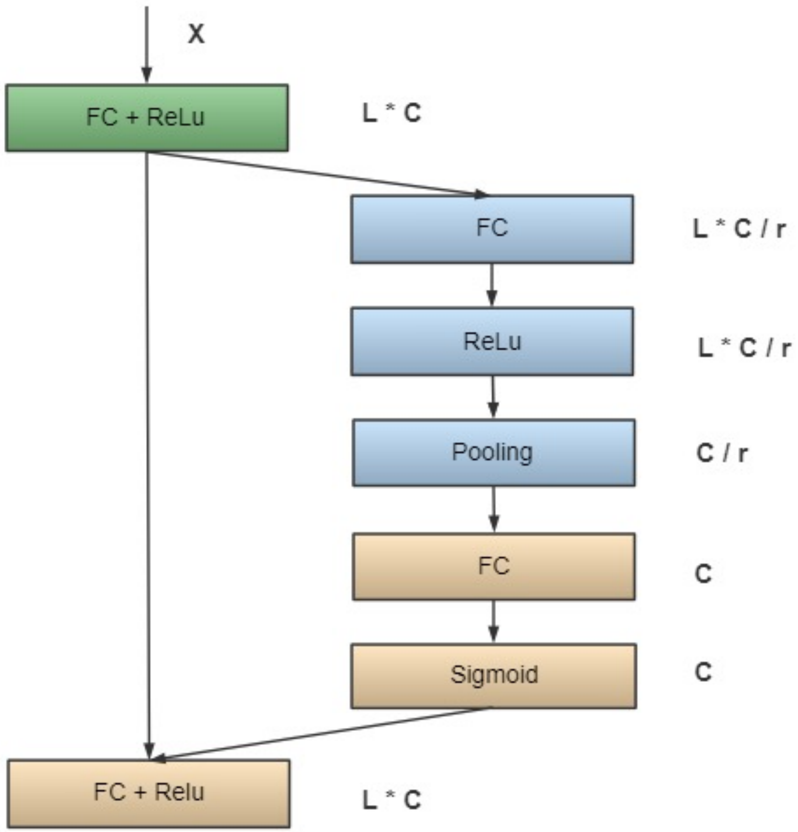
粗排
- 负采样
- 树模型

精排
- LambdaMart
- DNN升级
- TopK优化
- 多目标
- DeBias

重排
- **SE-Rank**
- **PRM**

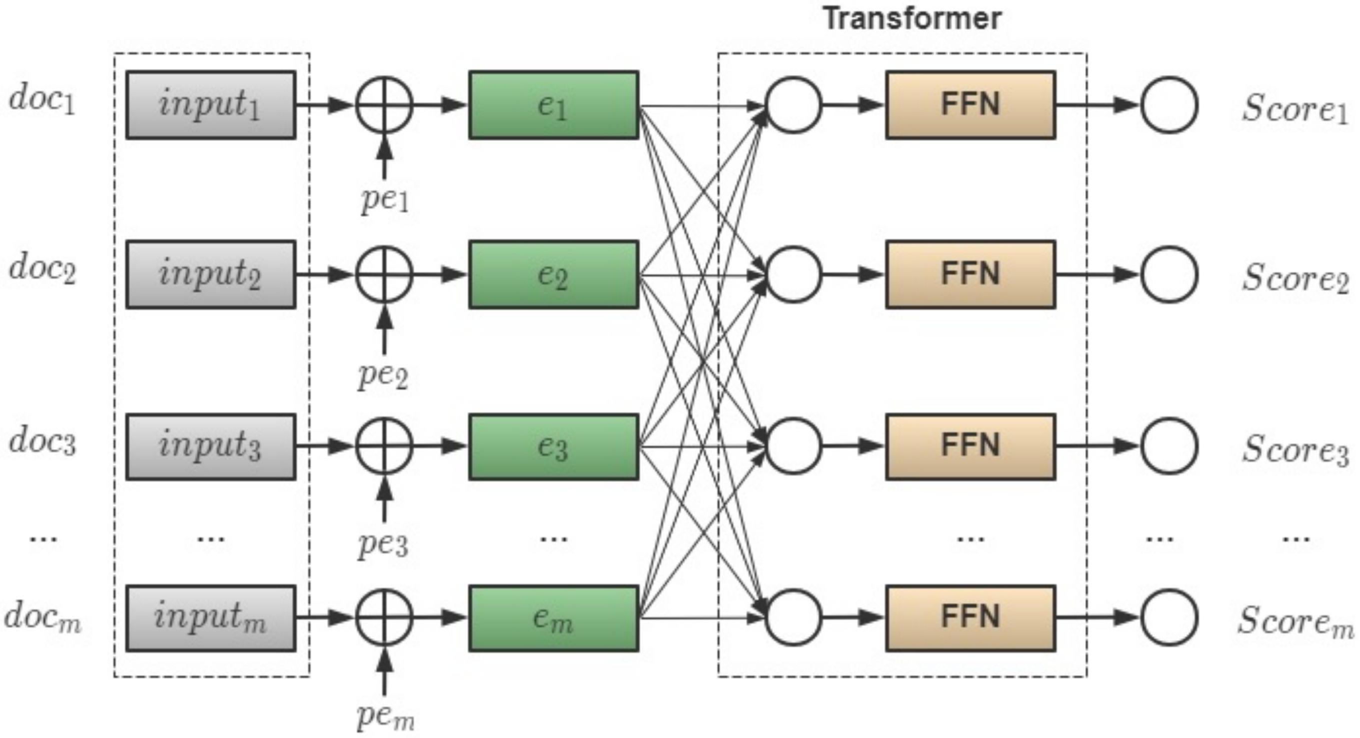
可解释性
- 稠密特征
- 稀疏特征

重排阶段 - SE-Rank



- 目标
  - 建模上下文感知
- 样本采样
  - List_size $\in [2, MaxLen]$
- 训练方式为List-wise
- 模型结构
  - 加入SE-Block
  - Field-wise

重排阶段 - PRM



- Listwise训练
- 模型结构
  - Doc特征表示
  - Transformer结构
- 离线评估
  - NDCG

提要

召回
- 倒排索引
- Bert召回

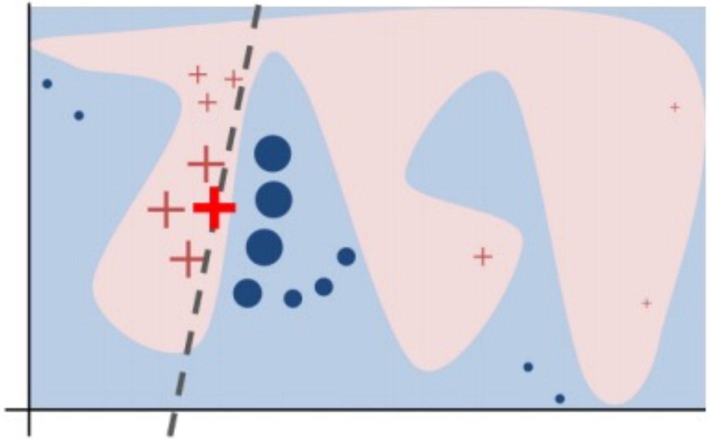
粗排
- 负采样
- 树模型

精排
- LambdaMart
- DNN升级
- TopK优化
- 多目标
- DeBias

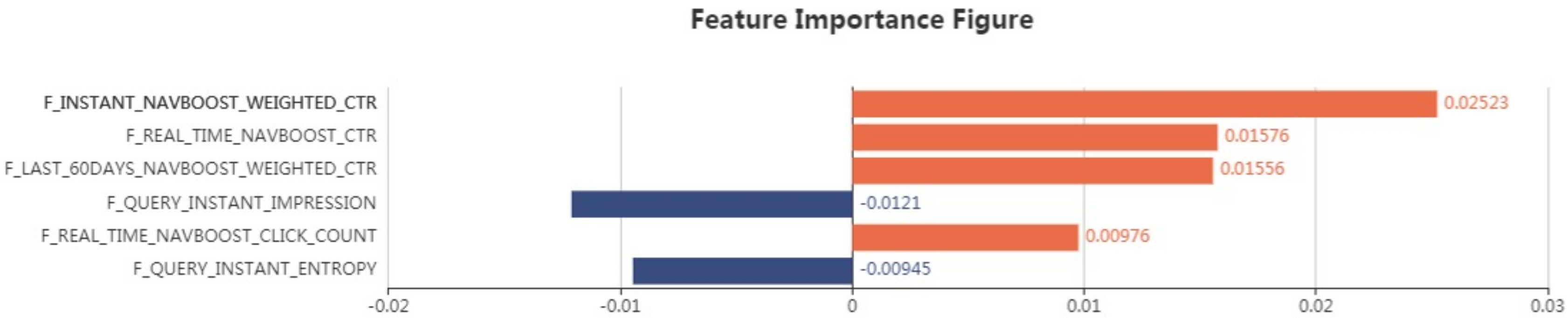
重排
- SE-Rank
- PRM

可解释性
- 稠密特征
- 稀疏特征

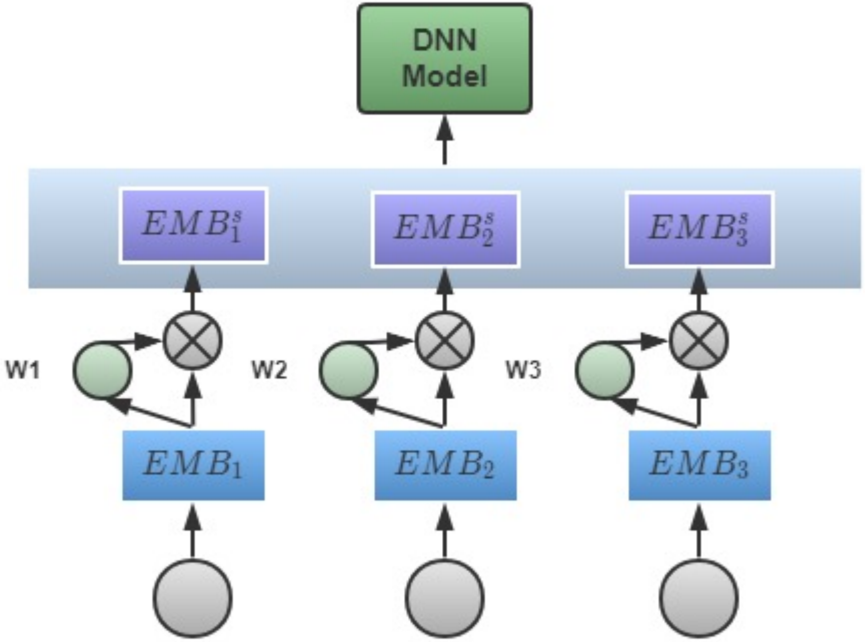
可解释性 - 稠密特征

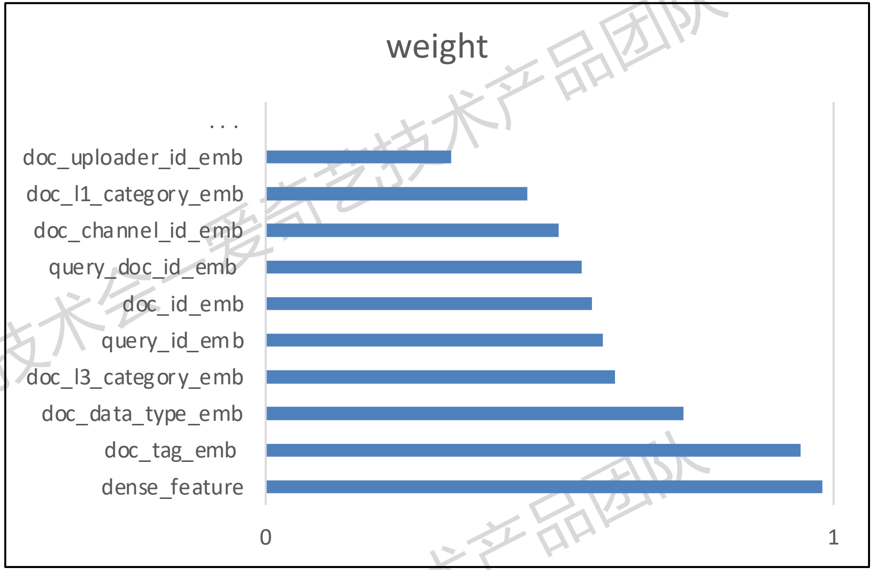


- ➢采用LIME框架
  - ➢局部可解释
- ➢特征填充
  - ➢稀疏特征填充为默认值
- ➢特征重要性
  - ➢实时交叉特征
  - ➢实时Query统计特征



可解释性 - 稀疏特征





- 全局可解释性
- 精排输入层加入简化版SE-Block
- 可解释性
  - Point-wise的SE-Block加入到精排有稀疏特征可解释的作用

Q&A

谢谢！

iQIYI 爱奇艺

悦 享 品 质