


# 涨指标的方法：特殊对待特殊人群

王树森

# 涨指标的方法有哪些？

- 改进召回模型，添加新的召回模型。
- 改进粗排和精排模型。
- 提升召回、粗排、精排中的多样性。
-  • 特殊对待新用户、低活用户等特殊人群。
- 利用关注、转发、评论这三种交互行为。

# 为什么要特殊对待特殊人群？

1. 新用户、低活用户的行为很少，个性化推荐不准确。
2. 新用户、低活用户容易流失，要想办法促使他们留存。
3. 特殊用户的行为（比如点击率、交互率）不同于主流用户，基于全体用户行为训练出的模型在特殊用户人群上有偏。

# 涨指标的方法

1. 构造特殊内容池，用于特殊用户人群的召回。
2. 使用特殊排序策略，保护特殊用户。
3. 使用特殊的排序模型，消除模型预估的偏差。

构造特殊的内容池

# 特殊内容池

- 为什么需要特殊内容池？
- 新用户、低活用户的行为很少，个性化召回不准确。（既然个性化不好，那么就保证内容质量好。）
- 针对特定人群的特点构造特殊内容池，提升用户满意度。  
（例如，对于喜欢留评论的中年女性，构造促评论内容池，满足这些用户的互动需求。）

# 如何构造特殊内容池

- 方法 1：根据物品获得的交互次数、交互率选择优质物品。
  - 圈定人群：只考虑特定人群，例如18~25岁一二线城市男性。
  - 构造内容池：用该人群对物品的交互次数、交互率给物品打分，选出分数最高的物品进入内容池。
  - 内容池有弱个性化的效果。
  - 内容池定期更新，加入新物品，排除交互率低和失去时效性的老物品。
  - 该内容池只对该人群生效。

# 如何构造特殊内容池

- 方法 1：根据物品获得的交互次数、交互率选择优质物品。
- 方法 2：做因果推断，判断物品对人群留存率的贡献，根据贡献值选物品。



# 特殊内容池的召回

- 通常使用双塔模型从特殊内容池中做召回。
  - 双塔模型是个性化的。
  - 对于新用户，双塔模型的个性化做不准。
  - 靠高质量内容、弱个性化做弥补。

# 特殊内容池的召回

- 通常使用双塔模型从特殊内容池中做召回。
- 额外的训练代价？
  - 对于正常用户，不论有多少内容池，只训练一个双塔模型。
  - 对于新用户，由于历史交互记录很少，需要单独训练模型。

# 特殊内容池的召回

- 通常使用双塔模型从特殊内容池中做召回。
- 额外的训练代价？
- 额外的推理代价？
  - 内容池定期更新，然后要更新 ANN 索引。
  - 线上做召回时，需要做 ANN 检索。
  - 特殊内容池都很小（比全量内容池小10~100倍），所以需要  
的额外算力不大。

# 特殊的排序策略

# 排除低质量物品

- 对于新用户、低活用户这样的特殊人群，业务上只关注留存，不在乎消费（总曝光量、广告收入、电商收入）。
- 对于新用户、低活用户，少出广告、甚至不出广告。
- 新发布的物品不在新用户、低活用户上做探索。
  - 物品新发布时，推荐做得不准，会损害用户体验。
  - 只在活跃的老用户上做探索，对新物品提权（boost）。
  - 不在新用户、低活用户上做探索，避免伤害用户体验。

# 差异化的融分公式

- 新用户、低活用户的点击、交互行为不同于正常用户。
- 低活用户的人均点击量很小；没有点击就不会有进一步的交互。
- 低活用户的融分公式中，提高预估点击率的权重（相较于普通用户）。
- 保留几个曝光坑位给预估点击率最高的几个物品。
  - 例：精排从 500 个物品中选 50 个作为推荐结果，其中 3 个坑位给点击率最高的物品，剩余 47 个坑位由融分公式决定。
  - 甚至可以把点击率最高的物品排在第一，确保用户一定能看到。

# 特殊的排序模型

# 差异化的排序模型

- 特殊用户人群的行为不同于普通用户。新用户、低活用户的点击率、交互率偏高或偏低。
- 排序模型被主流用户主导，对特殊用户做不准预估。
  - 用全体用户数据训练出的模型，给新用户做的预估有严重偏差。
  - 如果一个 APP 的用 90% 是女性，用全体用户数据训练出的模型，对男性用户做的预估有偏差。
- 问题：对于特殊用户，如何让排序模型预估做得准？



# 差异化的排序模型

- 方法 1：大模型 + 小模型。
  - 用全体用户行为训练大模型，大模型的预估  $p$  拟合用户行为  $y$ 。
  - 用特殊用户的行为训练小模型，小模型的预估  $q$  拟合大模型的残差  $y - p$ 。
  - 对主流用户只用大模型做预估  $p$ 。
  - 对特殊用户，结合大模型和小模型的预估  $p + q$ 。

# 差异化的排序模型

- 方法 1：大模型 + 小模型。
- 方法 2：融合多个 experts，类似 MMoE。
  - 只用一个模型，模型有多个 experts，各输出一个向量。
  - 对 experts 的输出做加权平均。
  - 根据用户特征计算权重。
  - 以新用户为例，模型将用户的新老、活跃度等特征作为输入，输出权重，用于对 experts 做加权平均。

# 差异化的排序模型

- 方法 1：大模型 + 小模型。
- 方法 2：融合多个 experts，类似 MMoE。
- 方法 3：大模型预估之后，用小模型做校准。
  - 用大模型预估点击率、交互率。
  - 将用户特征、大模型预估点击率和交互率作为小模型（例如 GBDT）的输入。
  - 在特殊用户人群的数据上训练小模型，小模型的输出拟合用户真实行为。

# 错误的做法

- 每个用户人群使用一个排序模型，推荐系统同时维护多个大模型。
  - 系统有一个主模型；每个用户人群有自己的一个模型。
  - 每天凌晨，用全体用户数据更新主模型。
  - 基于训练好的主模型，在某特殊用户人群的数据上再训练 1 epoch，作为该用户人群的模型。

# 错误的做法

- 每个用户人群使用一个排序模型，推荐系统同时维护多个大模型。
- 短期可以提升指标；维护代价大，长期有害。
  - 起初，低活男性用户模型比主模型的 AUC 高 0.2%。
  - 主模型迭代几个版本后，AUC 累计提升 0.5%。
  - 特殊人群模型太多，长期没有人维护和更新。
  - 如果把低活男性用户模型下线，换成主模型，在低活男性用户上的 AUC 反倒提升 0.3%！

# 总结：特殊对待特殊用户人群

- 召回：针对特殊用户人群，构造特殊的内容池，增加相应的召回通道。
- 排序策略：排除低质量物品，保护新用户和低活用户；特殊用户人群使用特殊的融分公式。
- 排序模型：结合大模型和小模型，小模型拟合大模型的残差；只用一个模型，模型有多个 experts；大模型预估之后，用小模型做校准。

**Thank You!**

<http://wangshusen.github.io/>