

涨指标的方法：提升多样性

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。

排序的多样性

精排多样性

- 精排阶段，结合兴趣分数和多样性分数对物品 i 排序。
 - s_i ：兴趣分数，即融合点击率等多个预估目标。
 - d_i ：多样性分数，即物品 i 与已经选中的物品的差异。
 - 用 $s_i + d_i$ 对物品做排序。

精排多样性

- 精排阶段，结合兴趣分数和多样性分数对物品 i 排序。
- 常用 MMR、DPP 等方法计算多样性分数，精排使用滑动窗口，粗排不使用滑动窗口。
 - 精排决定最终的曝光，曝光页面上邻近的物品相似度应该小。所以计算精排多样性要使用滑动窗口。
 - 粗排要考虑整体的多样性，而非一个滑动窗口中的多样性。

精排多样性

- 精排阶段，结合兴趣分数和多样性分数对物品 i 排序。
- 常用 MMR、DPP 等方法计算多样性分数，精排使用滑动窗口，粗排不使用滑动窗口。
- 除了多样性分数，精排还使用打散策略增加多样性。
 - 类目：当前选中物品 i ，之后 5 个位置不允许跟 i 的二级类目相同。
 - 多模态：事先计算物品多模态内容向量表征，将全库物品聚为 1000 类；在精排阶段，如果当前选中物品 i ，之后 10 个位置不允许跟 i 同属一个聚类。

粗排多样性

- 粗排给 5000 个物品打分，选出 500 个物品送入精排。
- 提升粗排和精排多样性都可以提升推荐系统核心指标。
- 根据 s_i 对 5000 个物品排序，分数最高的 200 个物品送入精排。
- 对于剩余的 4800 个物品，对每个物品 i 计算兴趣分数 s_i 和多样性分数 d_i 。
- 根据 $s_i + d_i$ 对剩余 4800 个物品排序，分数最高的 300 个物品送入精排。

召回的多样性

双塔模型： 添加噪声

- 用户塔将用户特征作为输入，输出用户的向量表征；然后做 ANN 检索，召回向量相似度高的物品。
- 线上做召回时（在计算出用户向量之后，在做 ANN 检索之前），往用户向量中添加随机噪声。
- 用户的兴趣越窄（比如用户最近交互的 n 个物品只覆盖少数几个类目），则添加的噪声越强。
- 添加噪声使得召回的物品更多样，可以提升推荐系统核心指标。

双塔模型：抽样用户行为序列

- 用户最近交互的 n 个物品（用户行为序列）是用户塔的输入。
- 保留最近的 r 个物品 ($r \ll n$) 。
- 从剩余的 $n - r$ 个物品中随机抽样 t 个物品 ($t \ll n$) 。（可以是均匀抽样，也可以用非均匀抽样让类目平衡。）
- 将得到的 $r + t$ 个物品作为用户行为序列，而不是用全部 n 个物品。
- 抽样用户行为序列为什么能涨指标？
 - 一方面，注入随机性，召回结果更多样化。
 - 另一方面， n 可以非常大，可以利用到用户很久之前的兴趣。

U2I2I: 抽样用户行为序列

- U2I2I (user \rightarrow item \rightarrow item) 中的第一个 item 是指用户最近交互的 n 个物品之一，在 U2I2I 中叫作种子物品。
- n 个物品覆盖的类目数较少，且类目不平衡。
 - 系统共有 200 个类目，某用户的 n 个物品只覆盖 15 个类目。
 - 足球类目的物品有 $0.4n$ 个，电视剧类目的物品有 $0.2n$ 个，其余类目的物品数均少于 $0.05n$ 个。

U2I2I: 抽样用户行为序列

- U2I2I ($\text{user} \rightarrow \text{item} \rightarrow \text{item}$) 中的第一个 item 是指用户最近交互的 n 个物品之一，在 U2I2I 中叫作种子物品。
- n 个物品覆盖的类目数较少，且类目不平衡。
- 做非均匀随机抽样，从 n 个物品中选出 t 个，让类目平衡。
(想法和效果与双塔中的用户行为序列抽样相似。)
- 用抽样得到的 t 个物品 (代替原本的 n 个物品) 作为 U2I2I 的种子物品。
- 一方面，类目更平衡，多样性更好。另一方面， n 可以更大，覆盖的类目更多。

探索流量

- 每个用户曝光的物品中有 2% 是非个性化的，用作兴趣探索。
- 维护一个精选内容池，其中物品均为交互率指标高的优质物品。（内容池可以分人群，比如30~40岁男性内容池。）
- 从精选内容池中随机抽样几个物品，跳过排序，直接插入最终排序结果。
- 兴趣探索在短期内负向影响核心指标，但长期会产生正向影响。

总结：提升多样性

- 精排：结合兴趣分数和多样性分数做排序；做规则打散。
- 粗排：只用兴趣分数选出部分物品；结合兴趣分数和多样性分数选出部分物品。
- 召回：往双塔模型的用户向量添加噪声；对用户行为序列做非均匀随机抽样（对双塔和 U2I2I 都适用）。
- 兴趣探索：保留少部分的流量给非个性化推荐。

Thank You!

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