

物品冷启动：评价指标

王树森

ShusenWang@xiaohongshu.com

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



物品冷启动

研究 UGC 的物品冷启

- 小红书上用户新发布的笔记。
- B站上用户新上传的视频。
- 今日头条上作者新发布的文章。

新笔记冷启动

为什么要特殊对待新笔记？

- 新笔记缺少与用户的交互，导致推荐的难度大、效果差。
- 扶持新发布、低曝光的笔记，可以增强作者发布意愿。

优化冷启的目标

1. 精准推荐：克服冷启的困难，把新笔记推荐给合适的用户，不引起用户反感。
2. 激励发布：流量向低曝光新笔记倾斜，激励作者发布。
3. 挖掘高潜：通过初期小流量的试探，找到高质量的笔记，给与流量倾斜。

评价指标

➡ • 作者侧指标：

- 发布渗透率、人均发布量。

➡ • 用户侧指标：

- 新笔记指标：新笔记的点击率、交互率。
- 大盘指标：消费时长、日活、月活。

➡ • 内容侧指标：

- 高热笔记占比。

作者侧指标

作者侧指标

发布渗透率 (penetration rate)

- 发布渗透率 = 当日发布人数 / 日活人数
- 发布一篇或以上，就算一个发布人数。
- 例：
 - 当日发布人数 = 100万
 - 日活人数 = 2000万
 - 发布渗透率 = $100 / 2000 = 5\%$

作者侧指标

人均发布量

- 人均发布量 = 当日发布笔记数 / 日活人数
- 例：
 - 每日发布笔记数 = 200万
 - 日活人数 = 2000万
 - 人均发布量 = $200 / 2000 = 0.1$

作者侧指标

- 发布渗透率、人均发布量反映出作者的发布积极性。
- 冷启的重要优化目标是促进发布，增大内容池。
- 新笔记获得的曝光越多，首次曝光和交互出现得越早，作者发布积极性越高。

用户侧指标

用户侧指标

新笔记的消费指标

- 新笔记的点击率、交互率。
 - 问题：曝光的基尼系数很大。
 - 少数头部新笔记占据了大部分的曝光。
- 分别考察高曝光、低曝光新笔记。
 - 高曝光：比如 >1000 次曝光。
 - 低曝光：比如 <1000 次曝光。

用户侧指标

大盘消费指标

- 大盘的消费时长、日活、月活。
- 大力扶持低曝光新笔记会发生什么？
 - 作者侧发布指标变好。
 - 用户侧大盘消费指标变差。

内容侧指标

内容侧指标

高热笔记占比

- 高热笔记：前 30 天获得 1000+ 次点击。
- 高热笔记占比越高，说明冷启阶段挖掘优质笔记的能力越强。

总结

- 作者侧指标：发布渗透率、人均发布量。
- 用户侧指标：新笔记消费指标、大盘消费指标。
- 内容侧指标：高热笔记占比。

冷启动的优化点

- 优化全链路（包括召回和排序）。
- 流量调控（流量怎么在新物品、老物品中分配）。

Thank You!

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

长期招聘优秀的算法工程师

- 部门：小红书社区技术部。
- 方向：搜索、推荐。
- 职位：校招、社招、实习。
- 地点：上海、北京。
- 联系方式：`ShusenWang@xiaohongshu.com`