

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）。

为什么给新笔记流量倾斜？

扶持新笔记的目的

- 目的1：促进发布，增大内容池。
 - 新笔记获得的曝光越多，作者创作积极性越高。
 - 反映在发布渗透率、人均发布量。
- 目的2：挖掘优质笔记。
 - 做探索，让每篇新笔记都能获得足够曝光。
 - 挖掘的能力反映在高热笔记占比。

工业界的做法

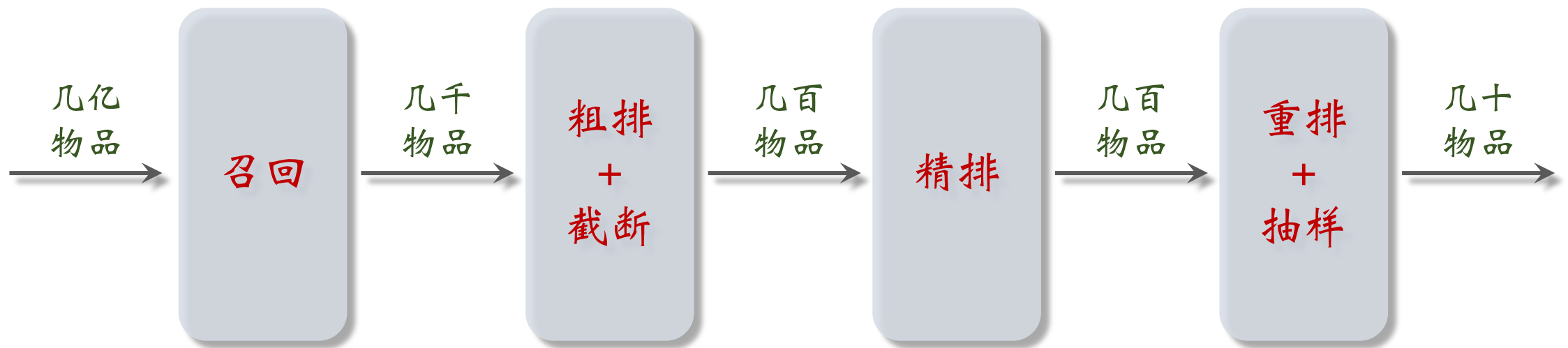
- 假设推荐系统只分发年龄 < 30 天的笔记。
- 假设采用自然分发，新笔记（年龄 < 24 小时）的曝光占比为 $1/30$ 。
- 扶持新笔记，让新笔记的曝光占比远大于 $1/30$ 。

流量调控技术的发展

1. 在推荐结果中强插新笔记。
2. 对新笔记的排序分数做提权 (boost) 。
3. 通过提权，对新笔记做保量。
4. 差异化保量。

新笔记提权 (boost)

推荐系统的链路



新笔记提权

- 目标：让新笔记有更多机会曝光。
 - 如果做自然分发，24小时新笔记占比为 $1/30$ 。
 - 做人为干涉，让新笔记占比大幅提升。
- 干涉粗排、重排环节，给新笔记提权。

新笔记提权

- 优点：容易实现，投入产出比好。
- 缺点：
 - 曝光量对提权系数很敏感。
 - 很难精确控制曝光量，容易过度曝光和不充分曝光。

新笔记保量

新笔记保量

- 保量：不论笔记质量高低，都保证 24 小时获得 100 次曝光。
- 在原有提权系数的基础上，乘以额外的提权的系数，比如：

		当前曝光次数			
		0~24次	25~49次	50~74次	75~100次
发布时间	0~5小时	1.0	1.0	1.0	1.0
	6~11小时	1.1	1.0	1.0	1.0
	12~17小时	1.2	1.1	1.0	1.0
	18~24小时	1.3	1.2	1.1	1.0

注：不是小红的真实数据

动态提权保量

用下面四个值计算提权系数

- ➡ • 目标时间：比如 24 小时。
- ➡ • 目标曝光：比如 100 次。
- ➡ • 发布时间：比如笔记已经发布 12 小时。
- ➡ • 已有曝光：比如笔记已经获得 20 次曝光。

$$\text{提权系数} = f\left(\frac{\text{发布时间}}{\text{目标时间}}, \frac{\text{已有曝光}}{\text{目标曝光}}\right) = f(\underline{0.5}, \underline{0.2})$$

保量的难点

保量成功率远低于 100%。

- 很多笔记在24小时达不到100次曝光。
- 召回、排序存在不足。
- 提权系数调得不好。

保量的难点

保量成功率远低于 100%。

线上环境变化会导致保量失败

- 线上环境变化：新增召回通道、升级排序模型、改变重排打散规则……
- 线上环境变化后，需要调整提权系数。

保量的难点

思考题

- 给所有新笔记一个很大的提权系数（比如 4 倍），直到达成 100 次曝光为止。
- 这样的保量成功率很高。
- 为什么不用这种方法呢？

保量的难点

给新笔记分数 boost 越多，对新笔记越有利？

- 好处：分数提升越多，曝光次数越多。
- 坏处：把笔记推荐给不太合适的受众。
 - 点击率、点赞率等指标会偏低。
 - 长期会受推荐系统打压，难以成长为热门笔记。

差异化保量

差异化保量

- **保量**：不论新笔记质量高低，都做扶持，在前 24 小时给 100 次曝光。
- **差异化保量**：不同笔记有不同保量目标，普通笔记保 100 次曝光，内容优质的笔记保 100~500 次曝光。

注：不是小红书真实数据

差异化保量

- 基础保量：24 小时 100 次曝光。
- 内容质量：用模型评价内容质量高低，给予额外保量目标，上限是加 200 次曝光。
- 作者质量：根据作者历史上的笔记质量，给予额外保量目标，上限是加 200 次曝光。
- 一篇笔记最少有 100 次保量，最多有 500 次保量。

注：不是小红书真实数据

总结

- ➡ • 流量调控：流量怎么在新老笔记之间分配。
- ➡ • 扶持新笔记：单独的召回通道、在排序阶段提权。
- ➡ • 保量：帮助新笔记在前 24 小时获得 100 次曝光。
- ➡ • 差异化保量：根据内容质量、作者质量，决定保量目标。