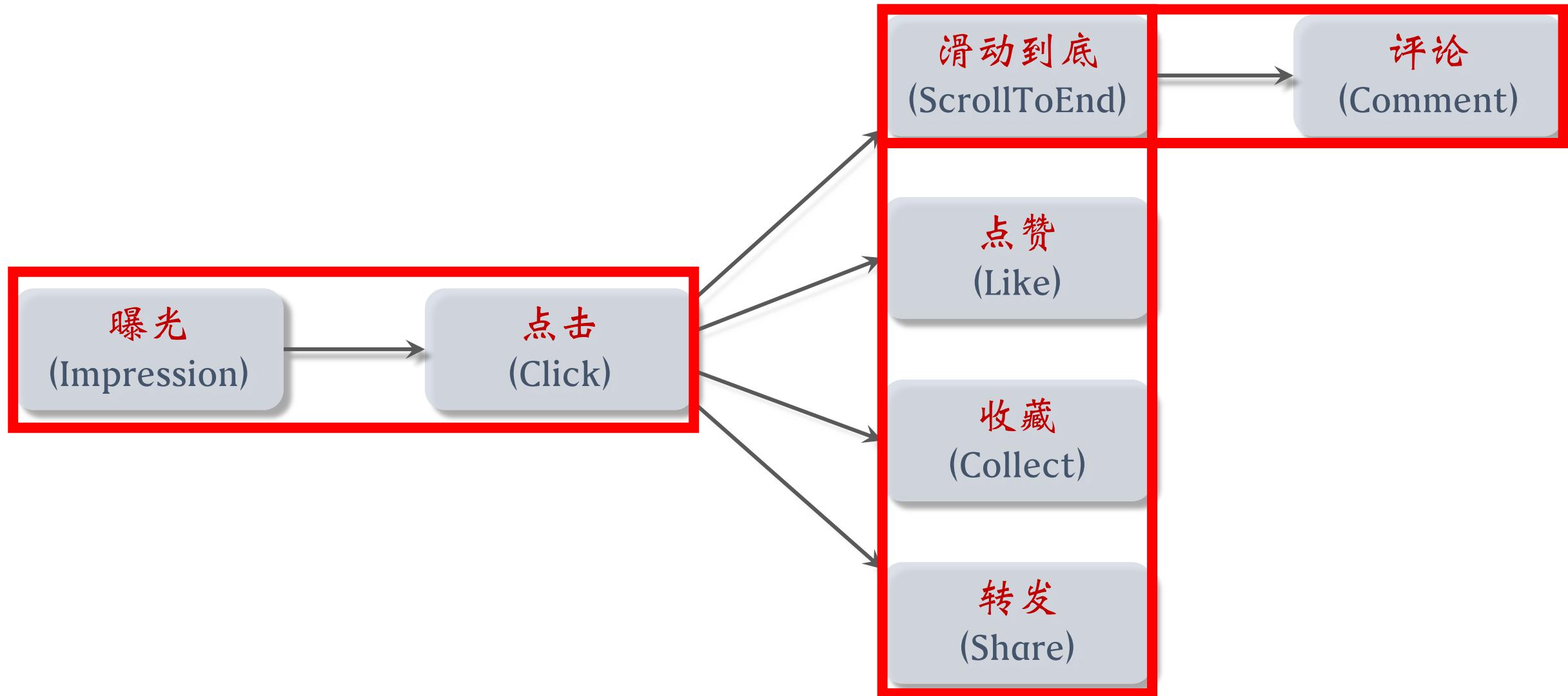


# 转化流程



# 消费指标

- 点击率 = 点击次数 / 曝光次数
- 点赞率 = 点赞次数 / 点击次数
- 收藏率 = 收藏次数 / 点击次数
- 转发率 = 转发次数 / 点击次数
- 阅读完成率 = 滑动到底次数 / 点击次数  $\times f(\text{笔记长度})$

# 北极星指标

→ • 用户规模：

- 日活用户数 (DAU) 、月活用户数 (MAU) 。

→ • 消费：

- 人均使用推荐的时长、人均阅读笔记的数量。

→ • 发布：

- 发布渗透率、人均发布量。

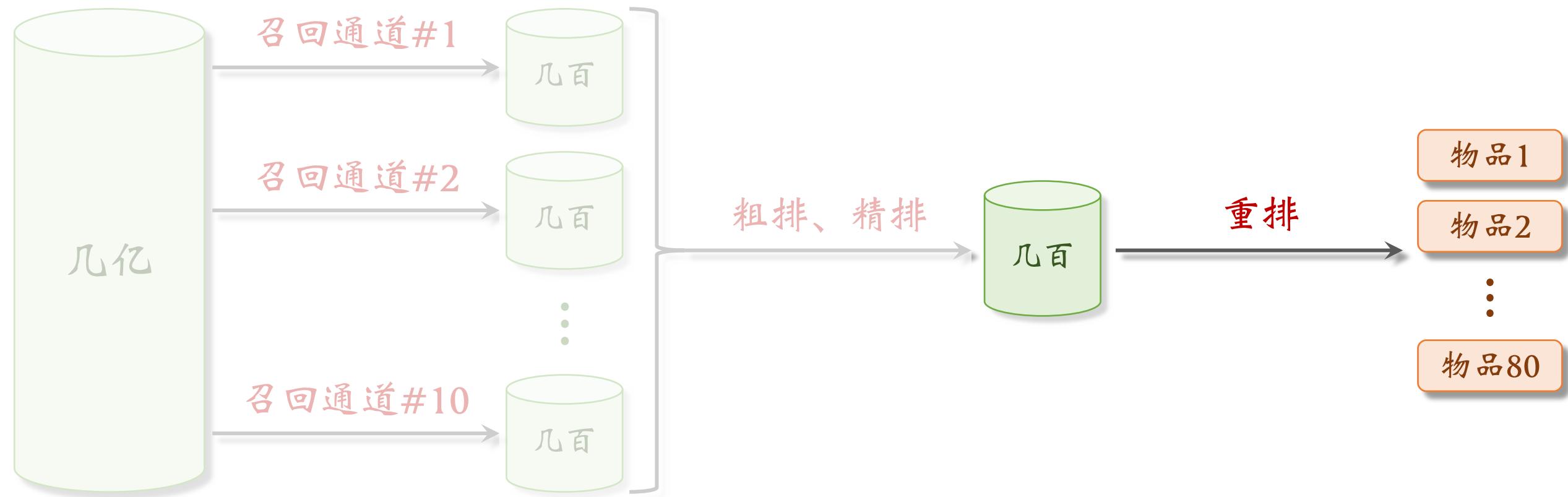
# 实验流程



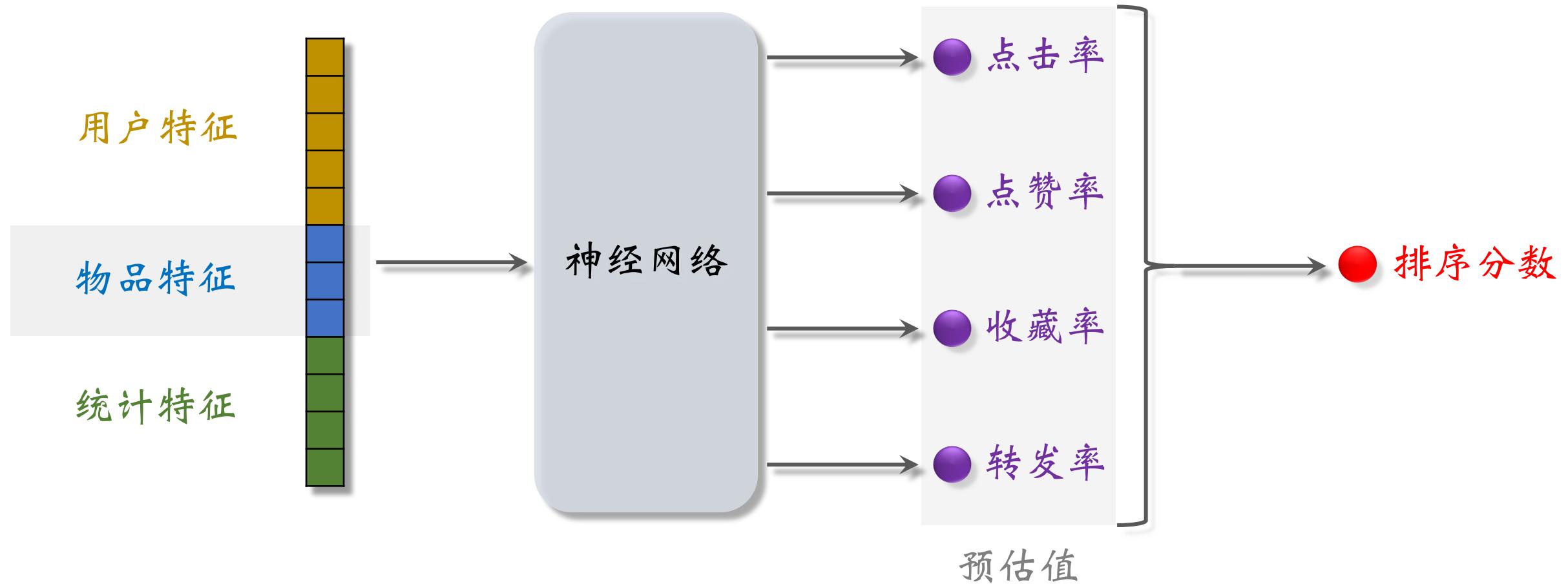
收集历史数据，在历史数据上做训练、测试。算法没有部署到产品中，没有跟用户交互。

把算法部署到实际产品中，用户实际跟算法做交互。

排序: [几千] → 粗排 → [几百] → 精排 → [几百] → 重排 → [几十]



## 粗排、精排



## 重排

- 做多样性抽样（比如MMR、DPP），从几百篇中选出几十篇。
- 用规则打散相似笔记。
- 插入广告、运营推广内容，根据生态要求调整排序。

# 总结

## 漏斗

- **召回**：用多条通道，取回几千篇笔记。
- **粗排**：用小规模神经网络，给几千篇笔记打分，选出分数最高的几百篇。
- **精排**：用大规模神经网络，给几百篇笔记打分。
- **重排**：做多样性抽样、规则打散、插入广告和运营笔记。

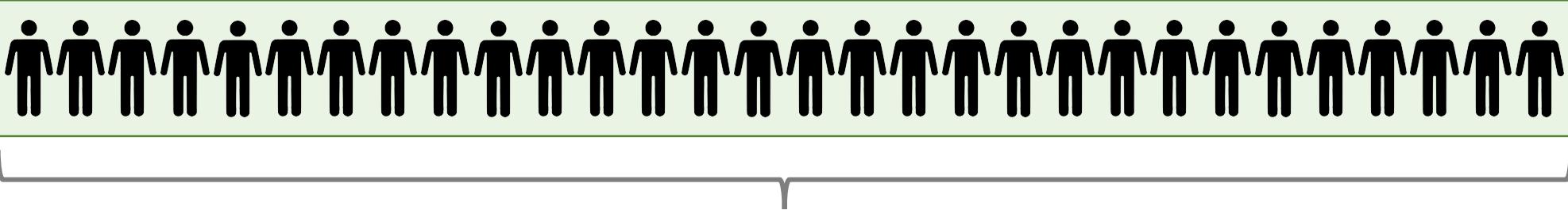
# 随机分桶



- 计算每个桶的业务指标，比如 DAU、人均使用推荐的时长、点击率、等等。
- 如果某个实验组指标显著优于对照组，则说明对应的策略有效，值得推全。

# 随机分桶

- 分  $b = 10$  个桶，每个桶中有 10% 的用户。
- 首先用哈希函数把用户 ID 映射成某个区间内的整数，然后把这些整数均匀随机分成  $b$  个桶。

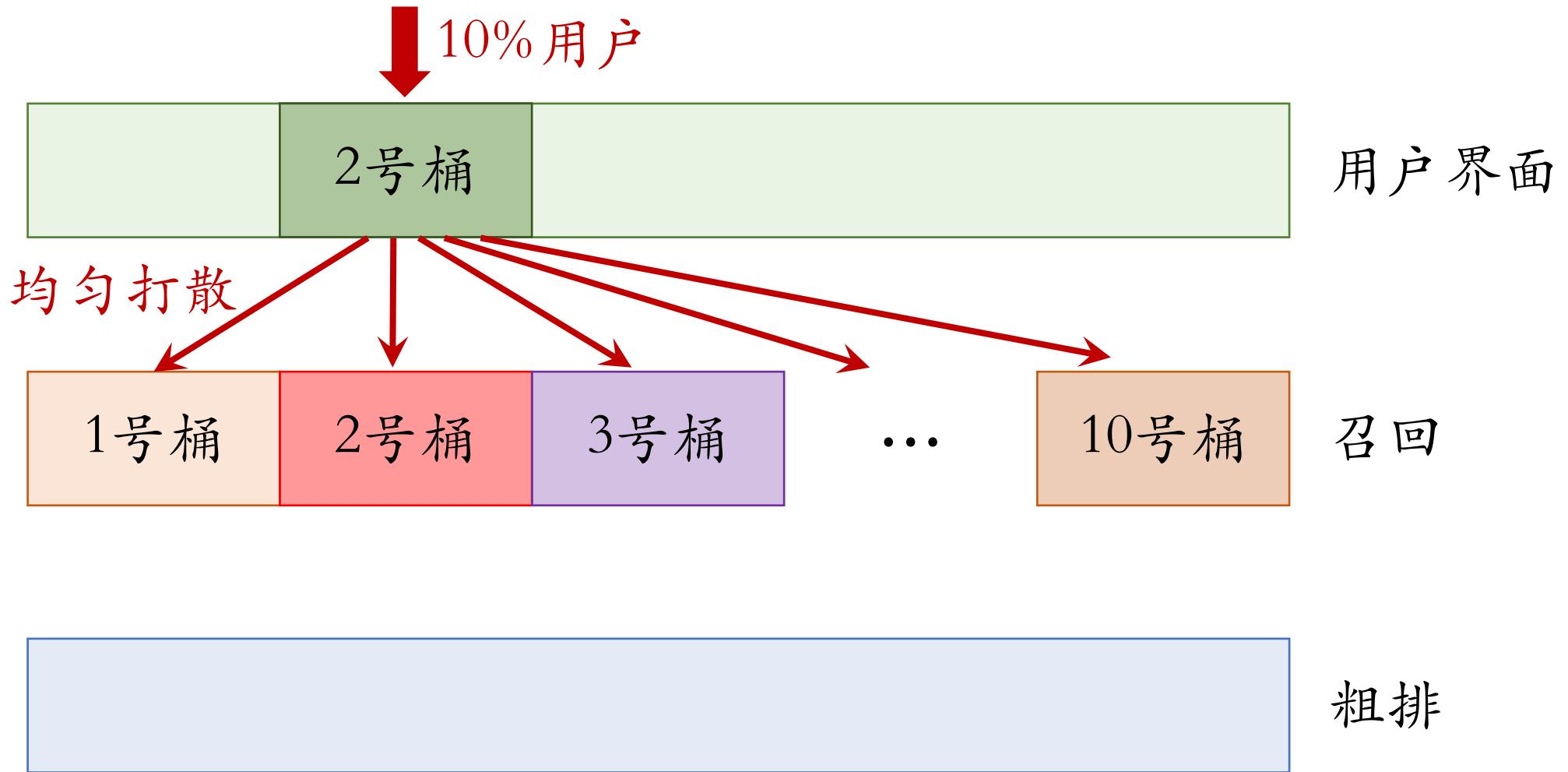


全部  $n$  位用户，分成  $b$  个桶，每个桶中有  $\frac{n}{b}$  位用户

# 分层实验

- 召回层把用户分成 10 个桶： $U_1, U_2, \dots, U_{10}$ 。
- 精排层把用户分成 10 个桶： $V_1, V_2, \dots, V_{10}$ 。
- 设系统共有  $n$  个用户，那么  $|U_i| = |V_j| = n/10$ 。
- 召回桶  $U_i$  和召回桶  $U_j$  交集为  $U_i \cap U_j = \emptyset$ 。
- 召回桶  $U_i$  和精排桶  $V_j$  交集的大小为  $|U_i \cap V_j| = n/100$ 。

# 不同层正交



# 互斥 vs 正交

- 如果所有实验都正交，则可以同时做无数组实验。
- 同类的策略（例如精排模型的两种结构）天然互斥，对于一个用户，只能用其中一种。
- 同类的策略（例如添加两条召回通道）效果会相互增强（ $1+1>2$ ）或相互抵消（ $1+1<2$ ）。互斥可以避免同类策略相互干扰。
- 不同类型的策略（例如添加召回通道、优化粗排模型）通常不会相互干扰（ $1+1=2$ ），可以作为正交的两层。

# Holdout 机制

- 每个实验（召回、粗排、精排、重排）独立汇报对业务指标的提升。
- 公司考察一个部门（比如推荐系统）在一段时间内对业务指标总体的提升。
- 取 10% 的用户作为 holdout 桶，推荐系统使用剩余 90% 的用户做实验，两者互斥。
- 10% holdout 桶 vs 90% 实验桶的 diff（需要归一化）为整个部门的业务指标收益。

# Holdout 机制

- 每个考核周期结束之后，清除 holdout 桶，让推全实验从 90% 用户扩大到 100% 用户。
- 重新随机划分用户，得到 holdout 桶和实验桶，开始下一轮考核周期。
- 新的 holdout 桶与实验桶各种业务指标的 diff 接近 0。
- 随着召回、粗排、精排、重排实验上线和推全，diff 会逐渐扩大。

# 反转实验

- 有的指标（点击、交互）立刻收到新策略影响，有的指标（留存）有滞后性，需要长期观测。
- 实验观测到显著收益后尽快推全新策略。目的是腾出桶供其他实验使用，或需要基于新策略做后续的开发。
- 用反转实验解决上述矛盾，既可以尽快推全，也可以长期观测实验指标。
- 在推全的新层中开一个旧策略的桶，长期观测实验指标。



# 总结

- **分层实验**：同层互斥（不允许两个实验同时影响一位用户）、不同层正交（实验有重叠的用户）。
- **Holdout**：保留 10% 的用户，完全不受实验影响，可以考察整个部门对业务指标的贡献。
- **实验推全**：新建一个推全层，与其他层正交。
- **反转实验**：在新的推全层上，保留一个小的反转桶，使用旧策略。长期观测新旧策略的 diff。