

物品冷启动：AB测试

王树森

ShusenWang@xiaohongshu.com

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

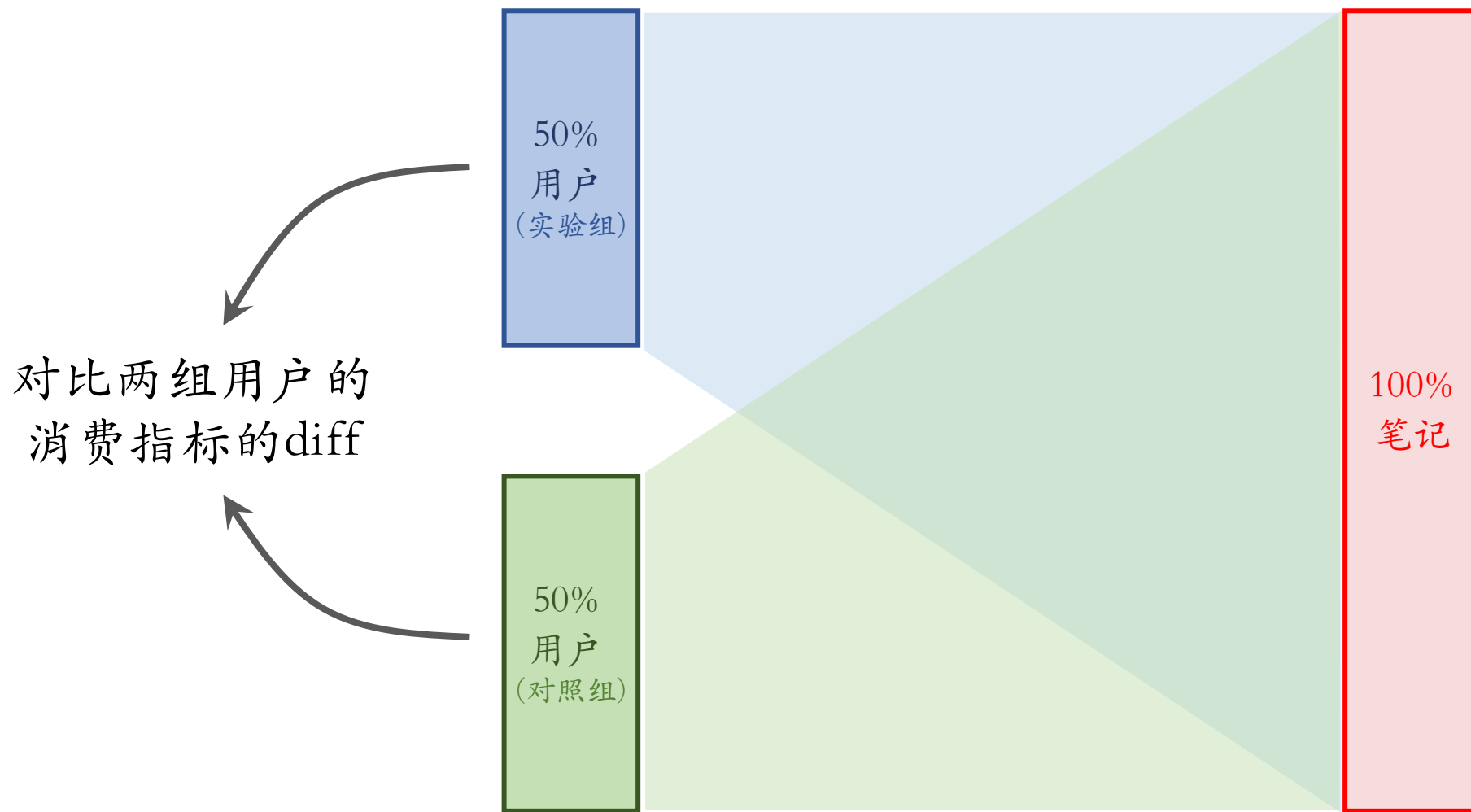


新笔记冷启的AB测试

- 作者侧指标：
 - 发布渗透率、人均发布量。
- 用户侧指标：
 - 对新笔记的点击率、交互率。
 - 大盘指标：消费时长、日活、月活。

标准的AB测试

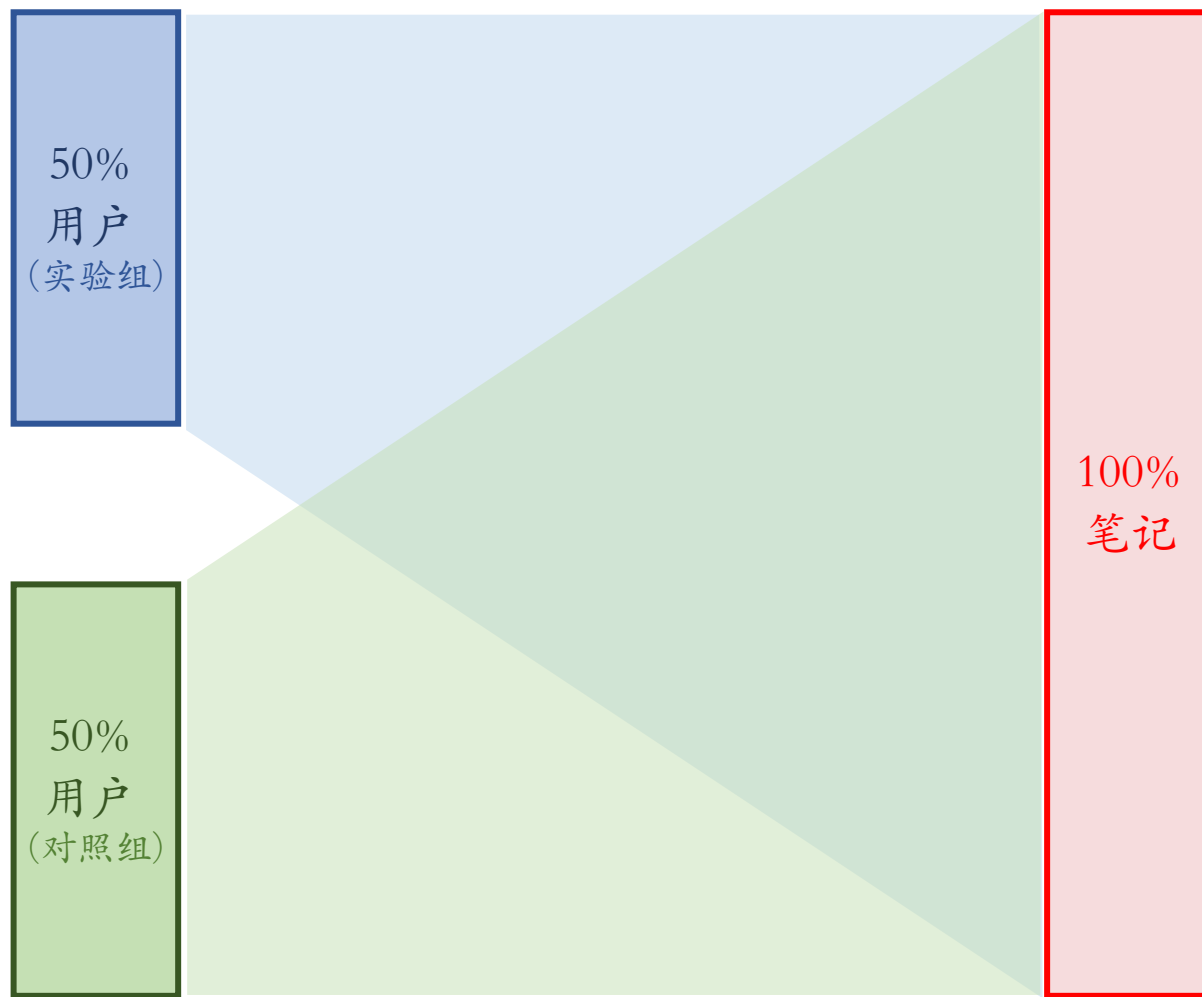
推荐系统标准的AB测试



用户侧实验

用户侧实验

用户侧



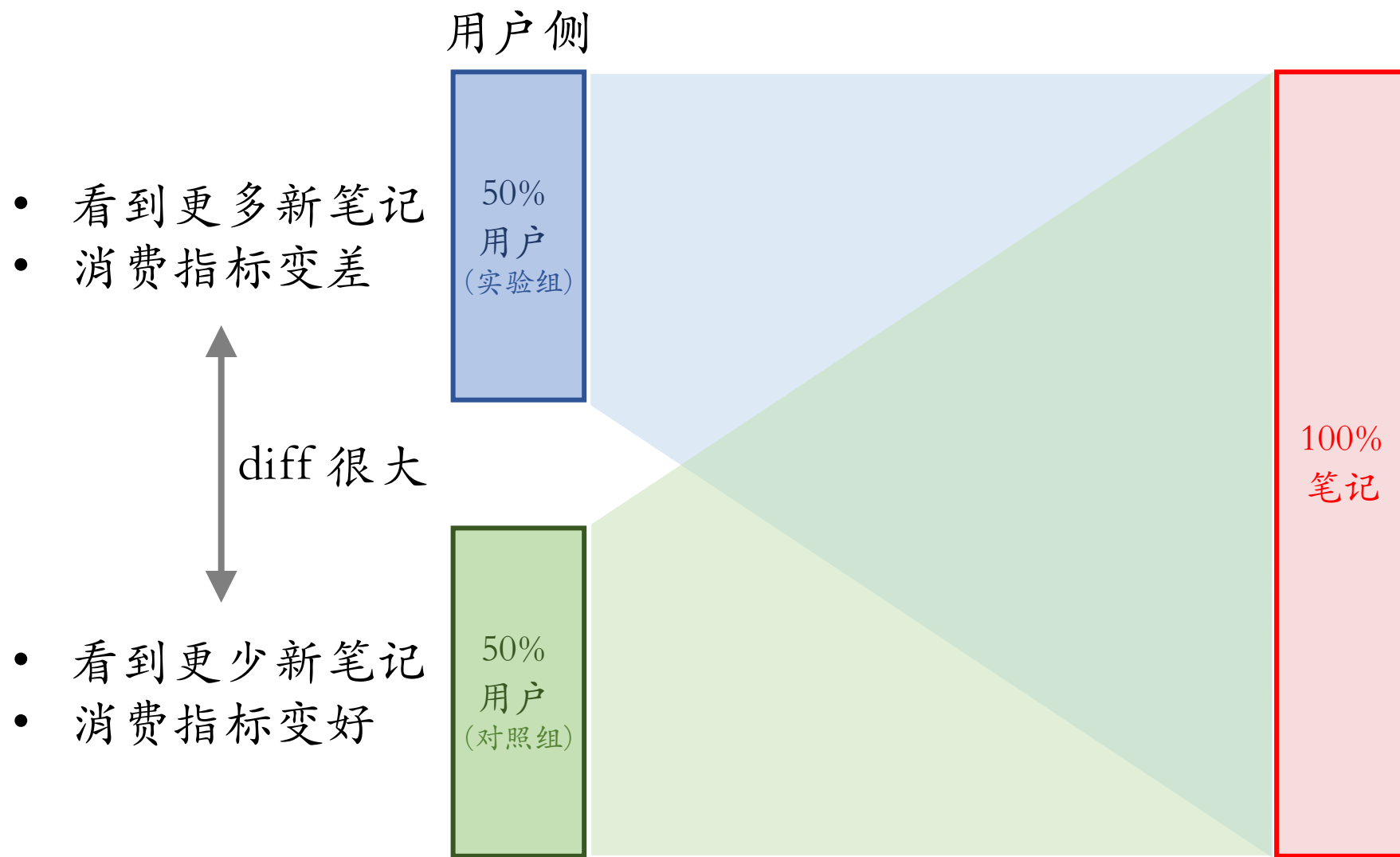
- 新笔记CTR的diff
- 用户消费时长的diff

用户侧实验

缺点

- 限定：保量 100 次曝光。
- 假设：新笔记曝光越多，用户使用APP时长越低。
- 新策略：把新笔记排序时的权重增大两倍。
- 结果（只看消费指标）：
 - AB测试的diff是负数（策略组不如对照组）。
 - 如果推全，diff会缩小（比如 $-2\% \rightarrow -1\%$ ）。

用户侧实验



作者侧实验

作者侧实验： 方案一

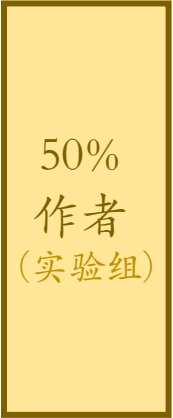
老笔记



用户侧



新笔记



作者侧实验： 方案一

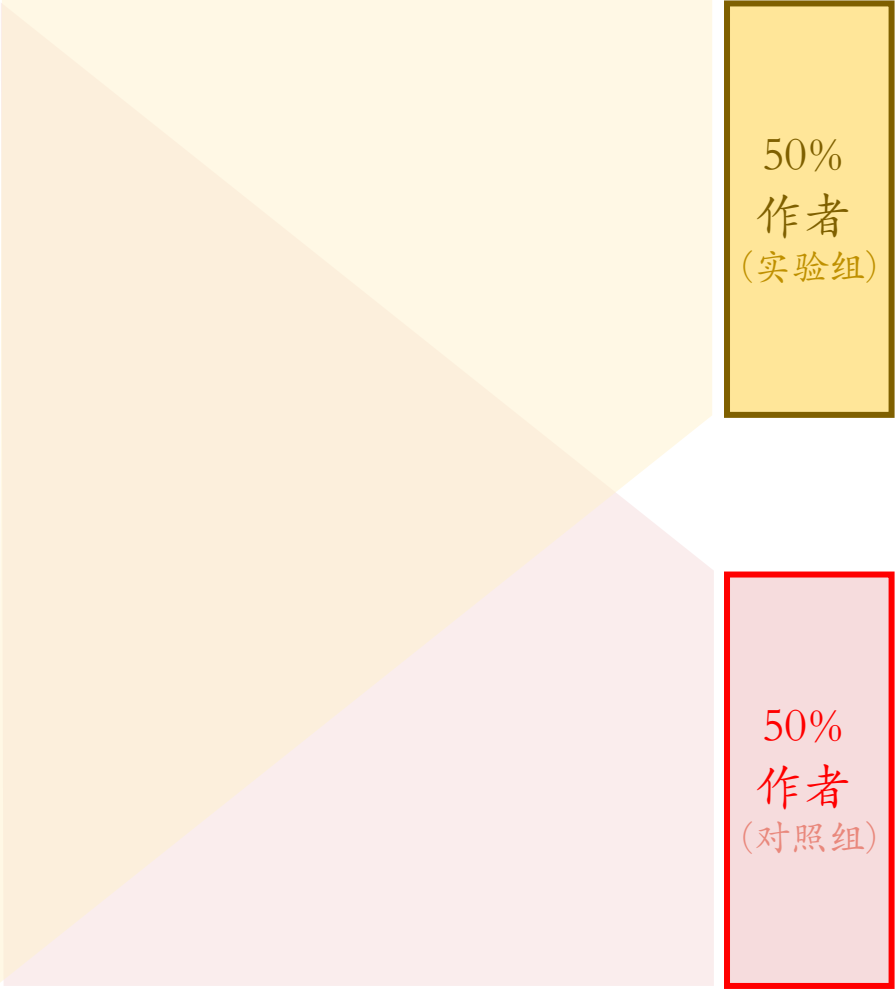
老笔记



用户侧



新笔记



新策略：
差异化保量

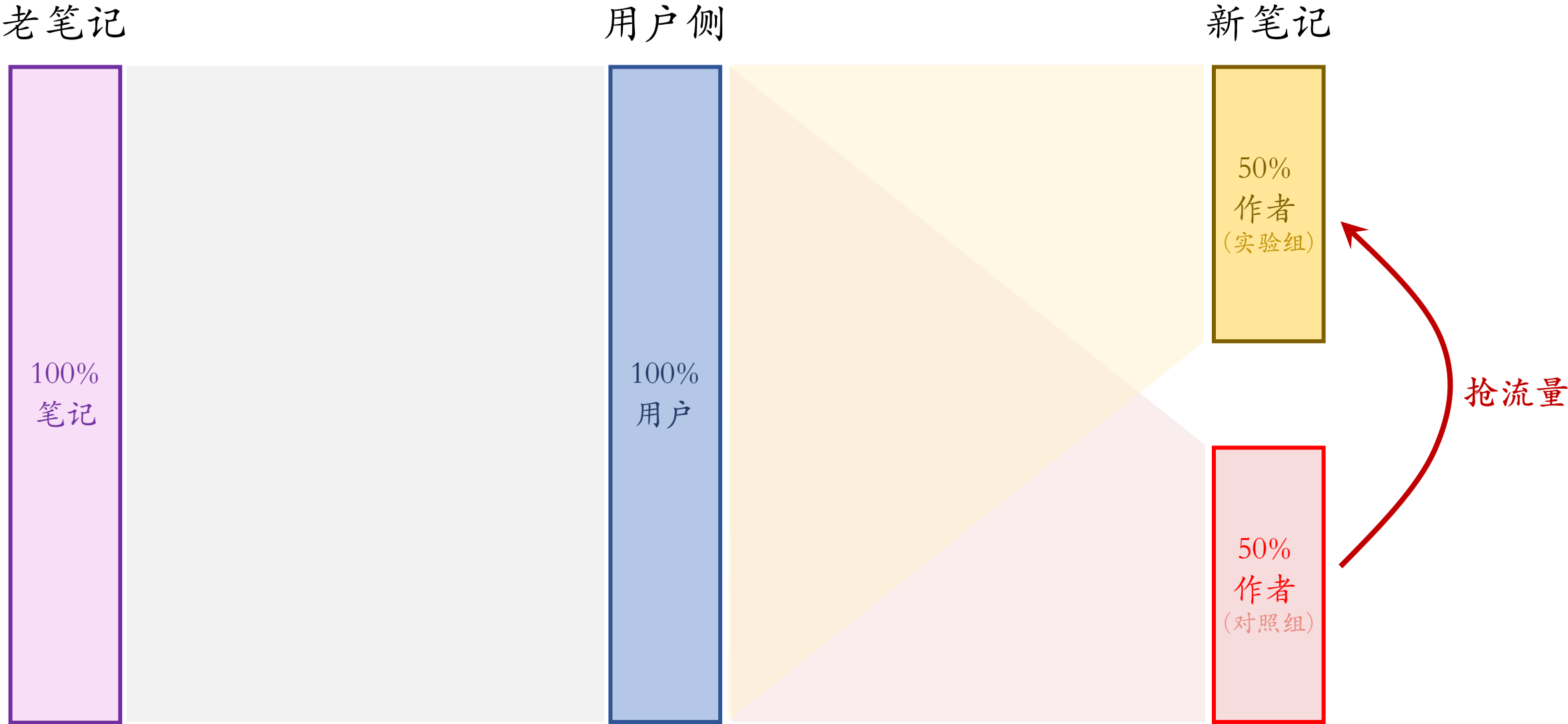
老策略：
简单的保量

作者侧实验：方案一

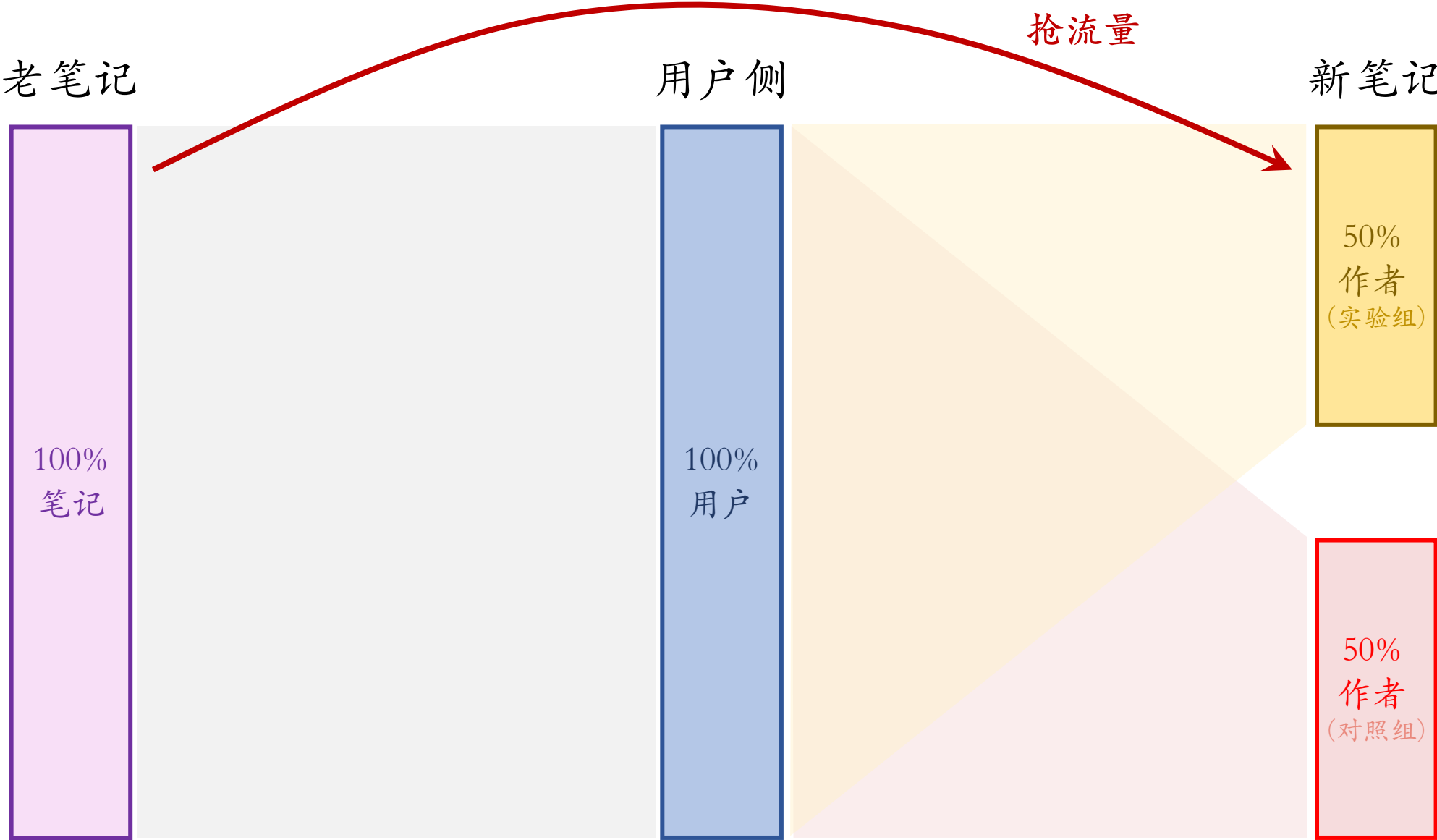
缺点：新笔记之间会抢流量

- 设定：
 - 新老笔记走各自队列，没有竞争。
 - 重排分给新笔记 $1/3$ 流量，分给老笔记 $2/3$ 流量
- 新策略：把新笔记的权重增大两倍。
- 结果（只看发布指标）：
 - AB测试的diff是正数（策略组优于对照组）。
 - 如果推全，diff会消失（比如 $2\% \rightarrow 0$ ）。

作者侧实验： 方案一



作者侧实验： 方案一

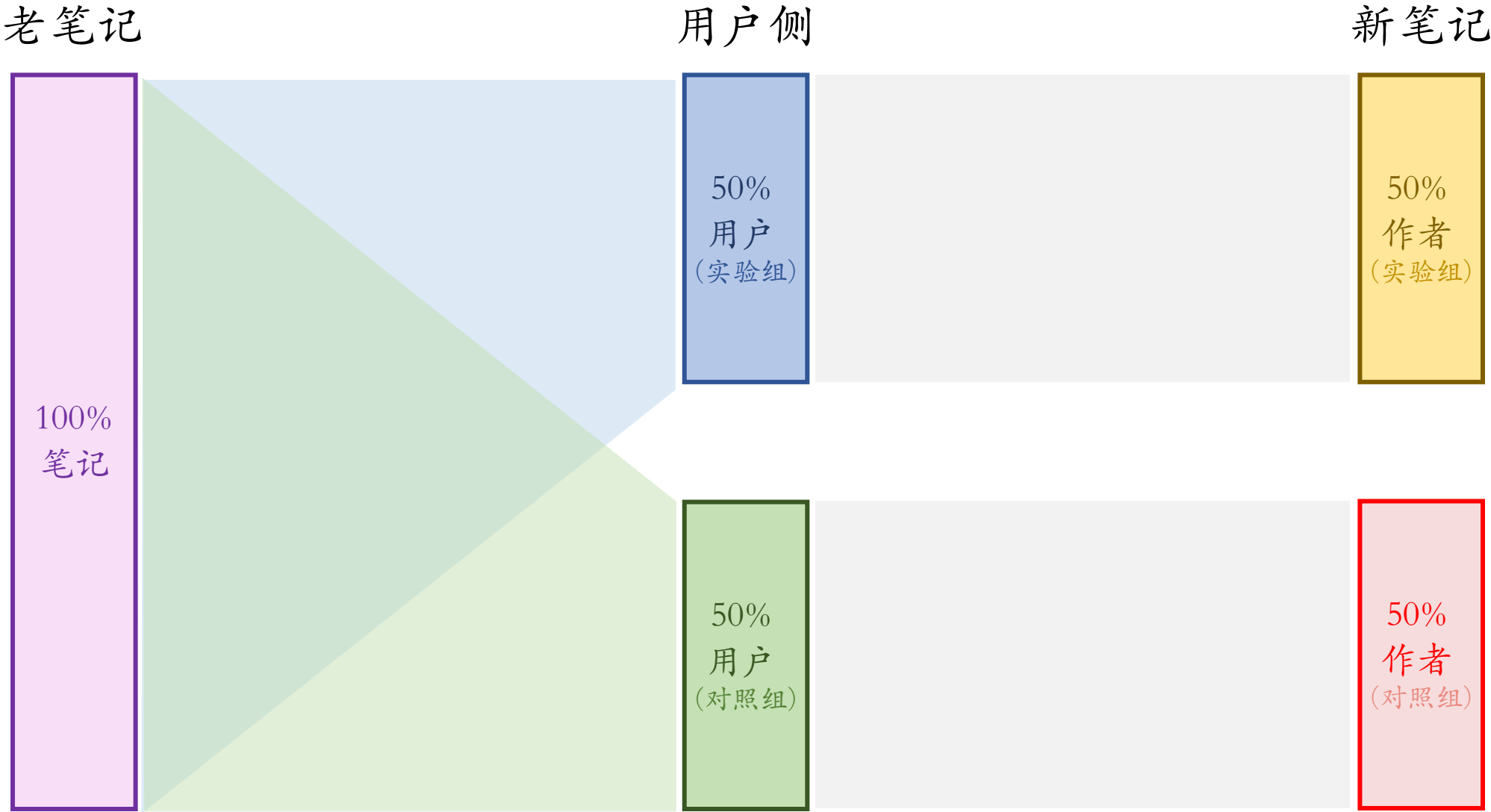


作者侧实验：方案一

缺点：新笔记和老笔记抢流量

- 设定：新老笔记自由竞争。
- 新策略：把新笔记排序时的权重增大两倍。
- AB 测试时，50%新笔记（带策略）跟 100%老笔记抢流量。
- 推全后，100%新笔记（带策略）跟 100%老笔记抢流量。
- 作者侧AB测试结果与推全结果有些差异。

作者侧实验： 方案二



作者侧实验：方案二

方案二比方案一的优缺点

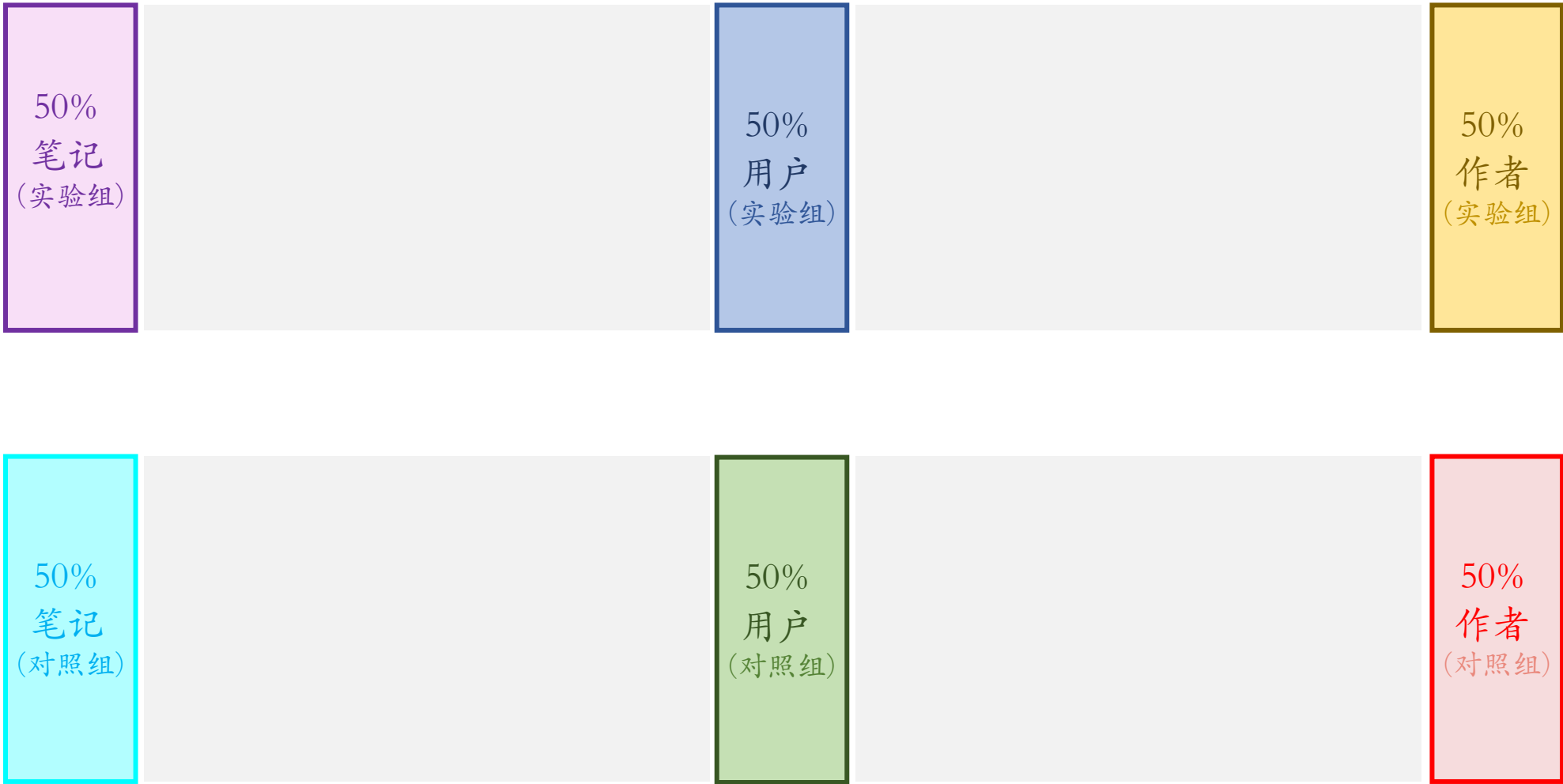
- 优点：新笔记的两个桶不抢流量，作者侧实验结果更可信。
- 相同：新笔记和老笔记抢流量，作者侧AB测试结果与推全结果有些差异。
- 缺点：新笔记池减小一半，对用户体验造成负面影响。

作者侧实验： 方案三

老笔记

用户侧

新笔记



总结

- 冷启的AB测试需要观测作者发布指标和用户消费指标。
- 各种AB测试的方案都有缺陷。（小红书有更好的方案，但也不完美。）
- 设计方案的时候，问自己几个问题：
 - 实验组、对照组新笔记会不会抢流量？
 - 新笔记、老笔记怎么抢流量？
 - 同时隔离笔记、用户，会不会让内容池变小？
 - 如果对新笔记做保量，会发生什么？

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`