

MTSC2019

中国移动互联网测试开发大会

Mobile Testing Summit China 2019

2019年6月28-29日 / 北京 国际会议中心

主办方: TesterHome  腾讯课堂

MTSC2019
中国移动互联网测试开发大会

爱奇艺推荐质量保证

段梅姣



短视频信息流

短视频信息流在近几年迅猛发展，成为互联网炙手可热的产品之一。



爱奇艺热点



爱奇艺极速版



推荐系统

几乎每一个信息流产品背后都有一个推荐系统，推荐系统帮助用户在海量的内容中寻找到感兴趣的内容。



姜饼



好多视频

01

视频推荐系统

主要组成要素简介

02

通用测试方法

痛点及挑战

03

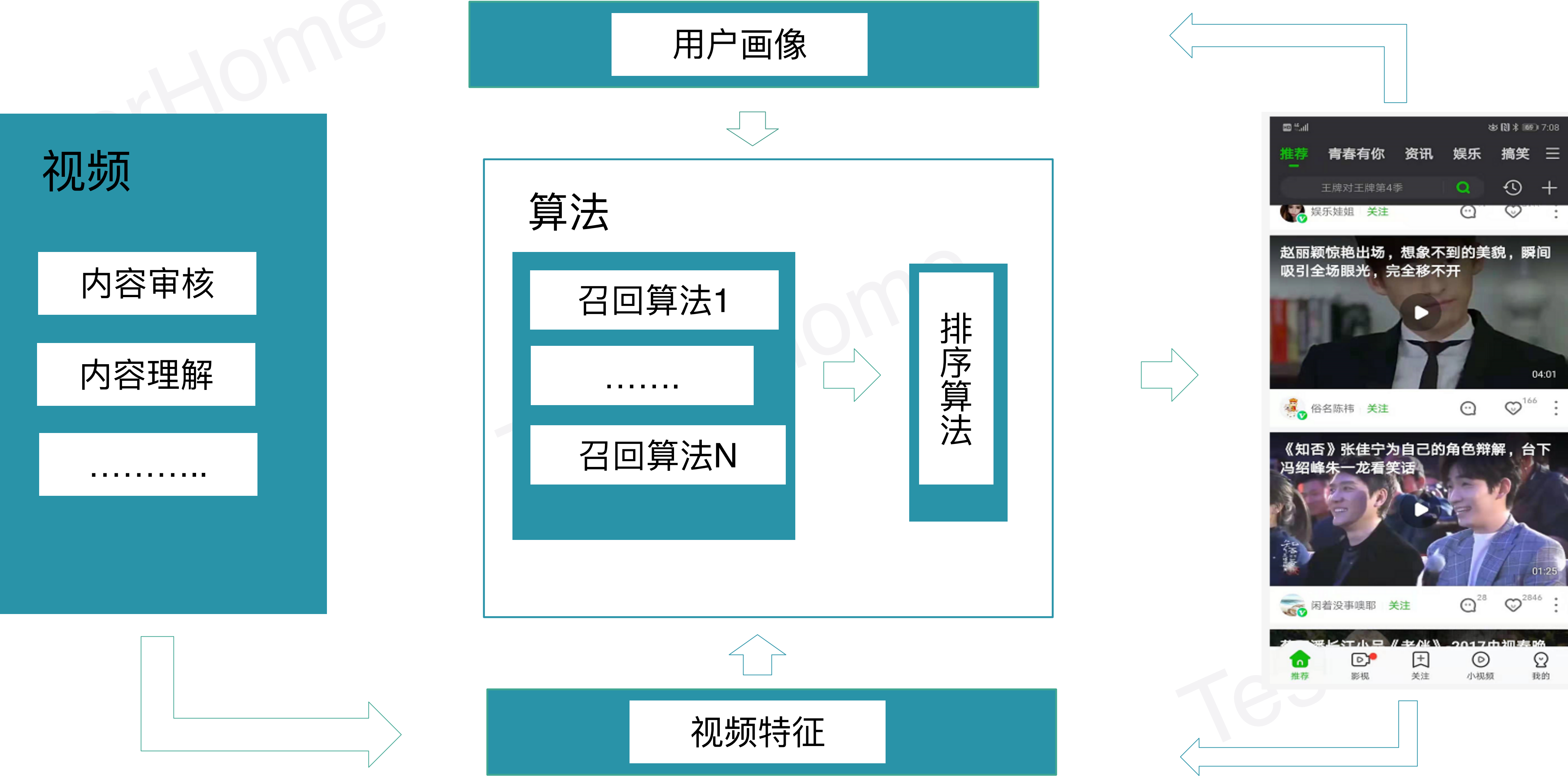
爱奇艺测试方法

线下测试目标和线上质量建设

04

总结

回顾



01 离线实验

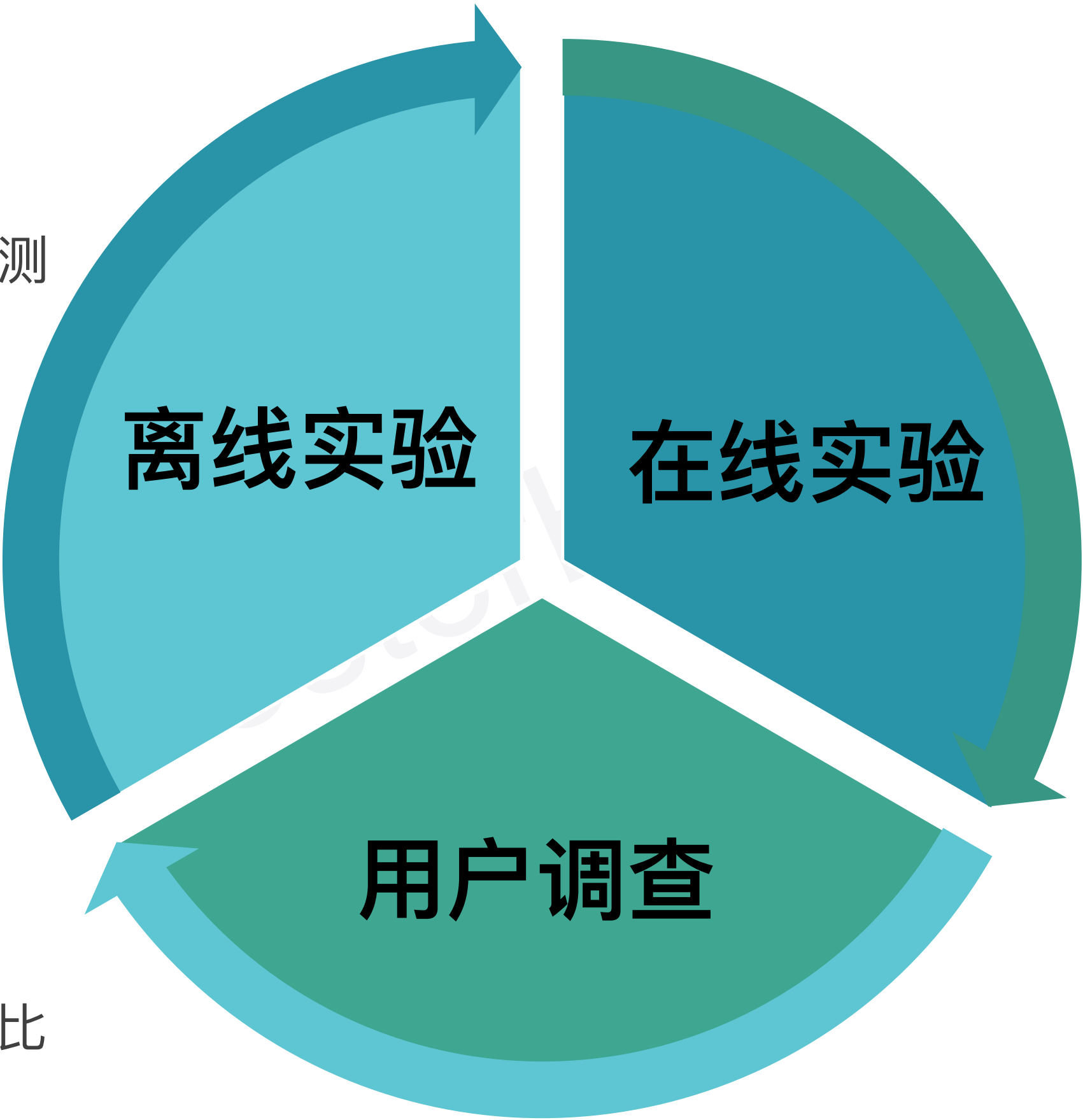
在离线数据集上通过离线指标预测
算法推荐效果

02 用户调查

通过真实用户给出主观反馈，
评估产品的体验

03 在线实验

能评估产品指标（商业指标），比
如人均点击、次日留存等



通用难点

- 1， 离线实验无法准确评估对产品线上指标的影响
- 2， 用户调查成本高、实施难
- 3， 在线实验周期长，反馈模糊



挑战一

如何把真实感受量化为指标



挑战二

如何在线上快速收集反馈并聚焦
发现问题

线下
测试

小流量
实验

线上
质量

快速交付

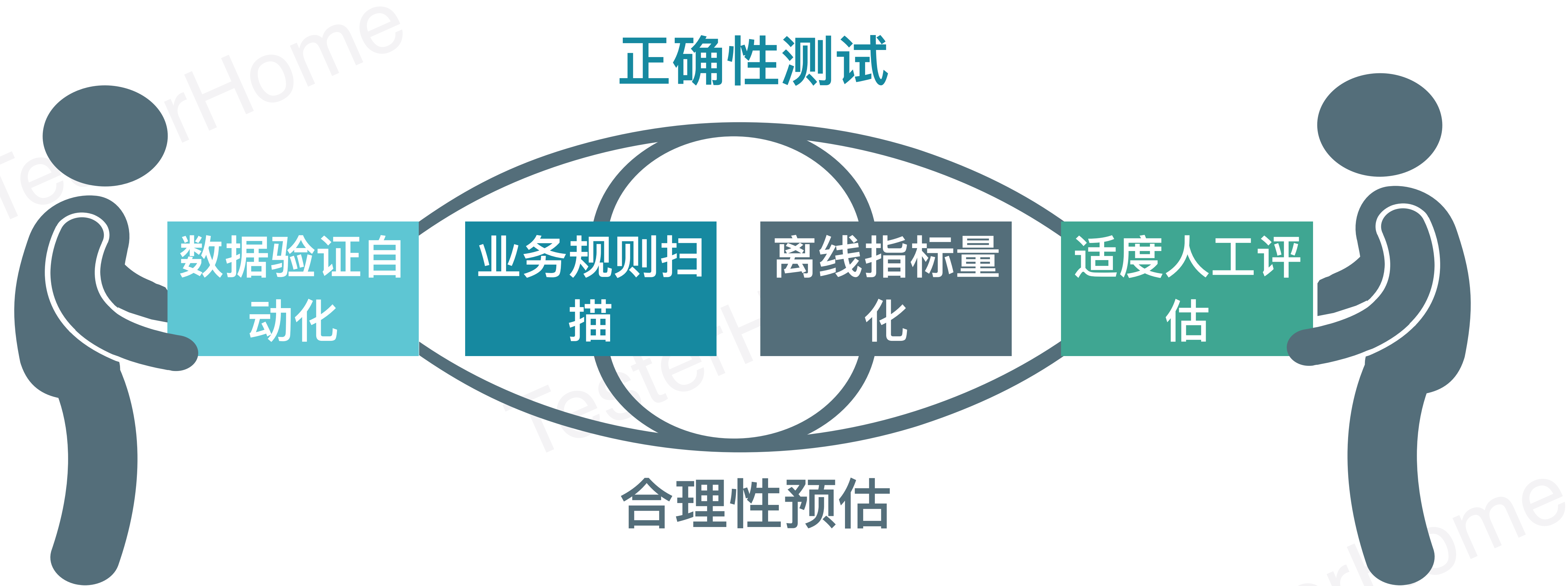
重点保证规则正确性；
初步评估上线后的影响；

效果指标统计及分析

产品指标横向对比评估效果；
结果回溯定性分析聚焦问题；

质量监控

监控及时发现问题；
定期评测产品体验；



正确性测试保证没有严重功能及体验性问题；
合理性预估保证不会导致核心指标明显负面波动。

数据

自动化校验

字段正确性

字段完整性

数量

.....

小时级返回结果

业务

规则扫描

过滤规则

内容无重复

封面图无重复

.....

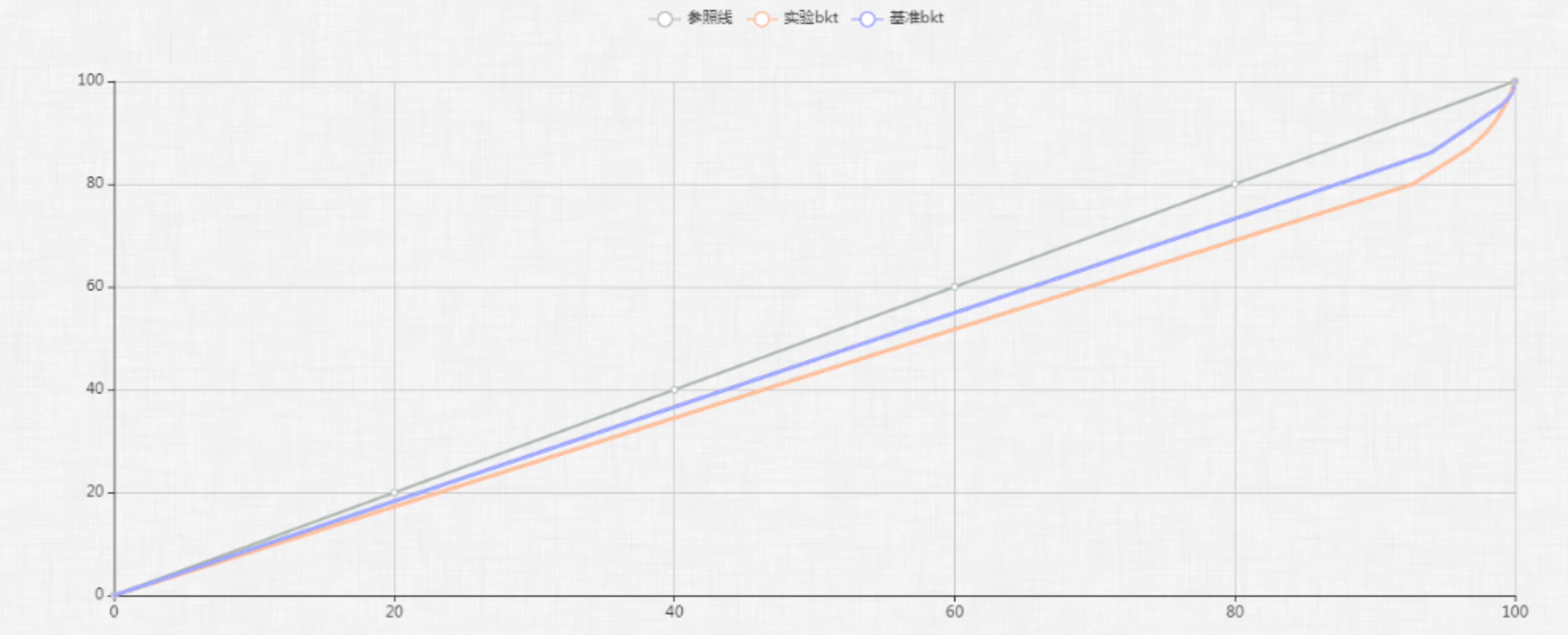
分钟级返回结果



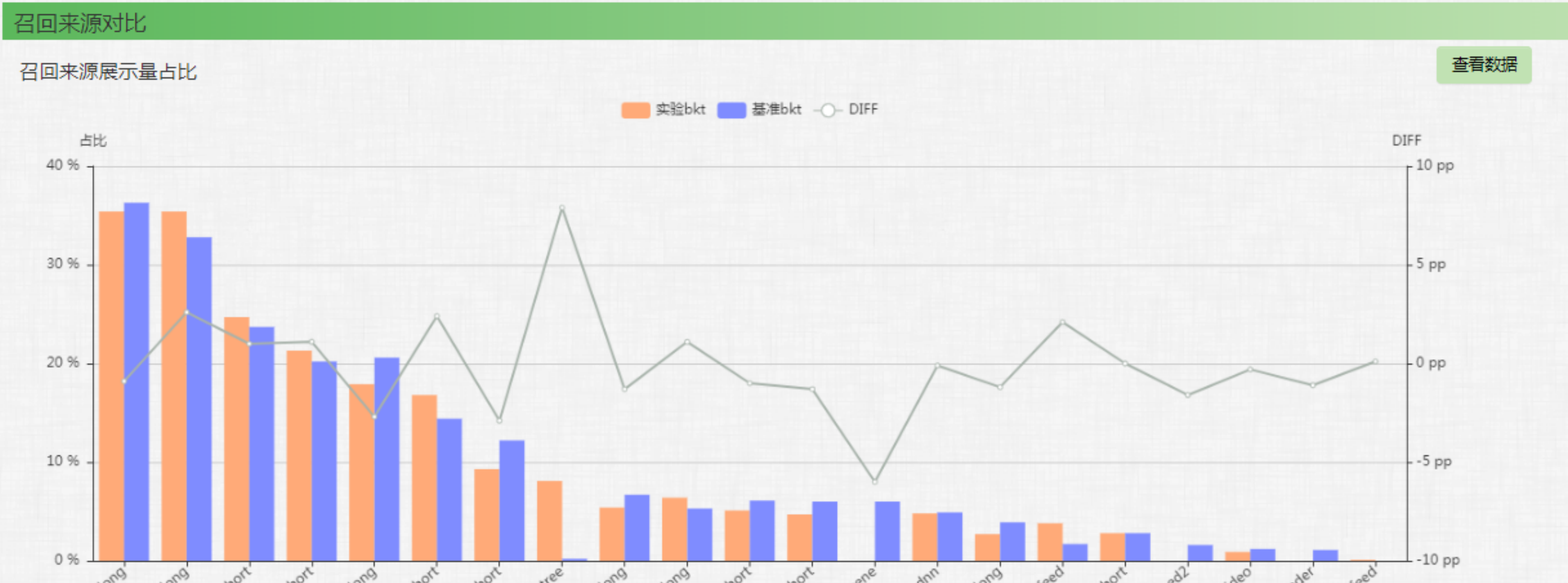
基尼系数结果对比

实验bkt	基准bkt	DIFF
0.081	0.131	-0.050

基尼系数对比曲线图



快速导航：[召回来源对比](#) [分发效果对比](#) [feed内容对比](#) [排序结果对比](#)





定量分析效果

横向对比效果指标，通过指标波动衡量效果：展示量、点击量、CTR、UCTR、次日留存、人均展示时长、人均播放时长、等

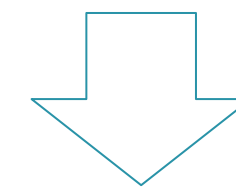


定性佐证结论

还原视频推出全过程：从视频特征+用户画像->召回算法->用户反馈，抽样分析聚焦问题

有了定量为什么还要定性分析？

- 1，效果指标有时升有时降，为什么？
- 2，指标有的升有的降，该不该全量？
- 3，效果指标比预期有差距，原因是？
- 4，时间短无法评估用户侧影响，怎么办？

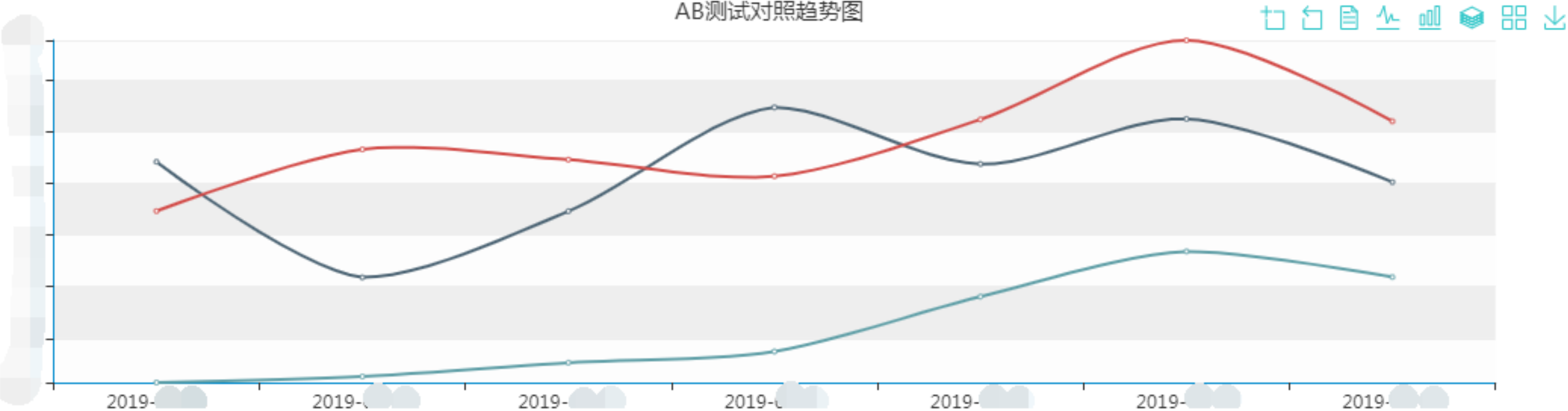


定性分析帮助缩短周期，更聚焦发现问题

- 效果指标
- ☒ 展示人均时长(分钟)
- ☒ 点击人均时长(分钟)
- ☐ 展均播放时长(秒)
- ☐ CTR
- ☐ UCTR
- ☐ 人均展示视频数
- ☐ 次日留存率(%)
- ☐ 点击次日留存率(%)
- ☐ 推出视频平均时长(分钟)
- ☐ 推出视频总时长(分钟)

- 基础指标
- ☐ 点击量
- ☐ 展示量
- ☐ 展示UV
- ☐ 点击UV
- ☐ 时长(分钟)

AB测试对照趋势图



01.样本数据

根据定量指标结果，抽取用户样本

02.回溯过程

还原内容推出过程：
包括用户画像、推荐结果、推荐原因、排序结果、用户行为

03.对比分析

从过程及用户实际行为反馈对比正负样本、基准样本

04.聚焦问题

总结结果，聚焦发现问题，并辅以通用指标评估结果

定性分析帮助聚焦问题

通用指标：准确性，多样性，惊喜度，实时性；

建立线上监控和评测体系





用户：能更快的找到更多喜欢的内容

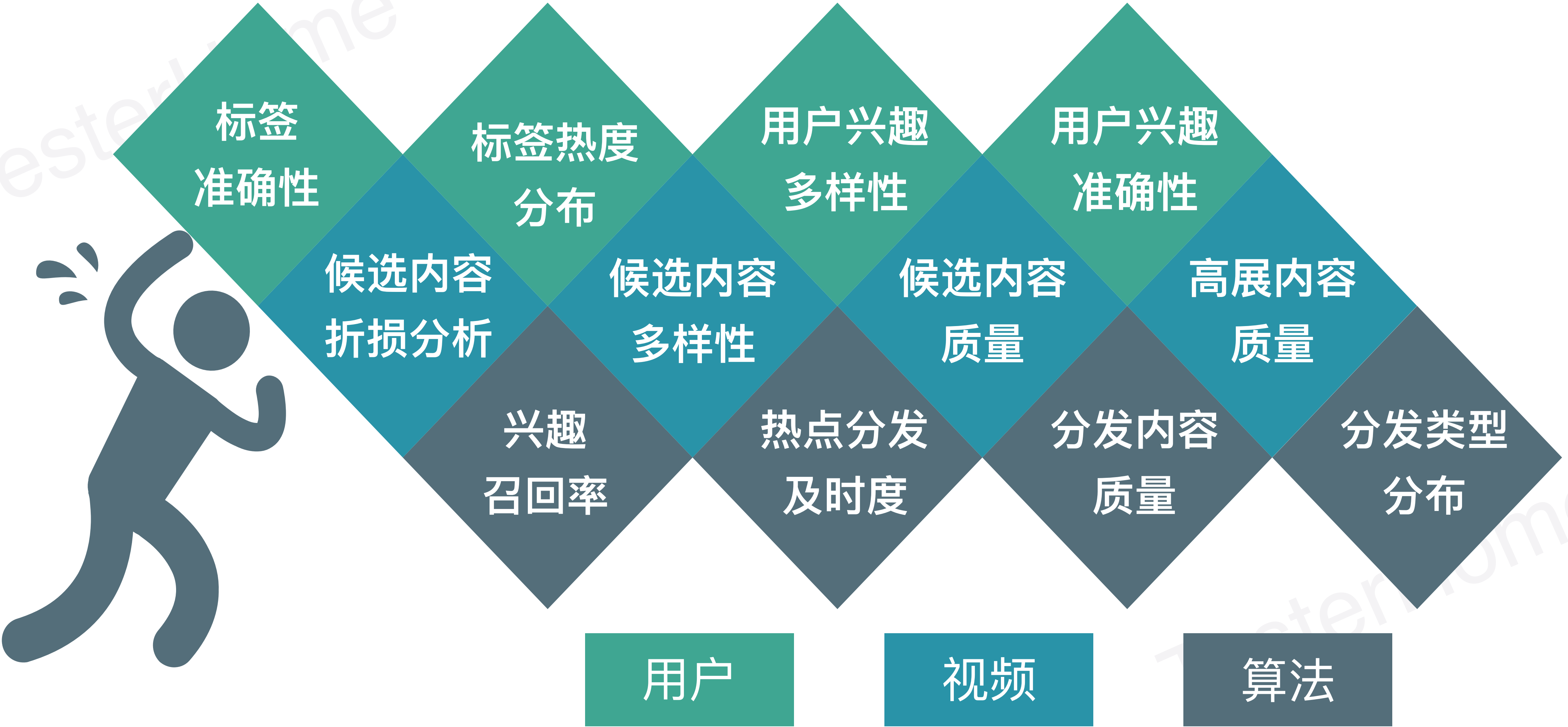
从进入平台到找到自己想要的内容的时间间隔；
在M篇内容内喜欢的内容越多越好；
新用户、活跃用户、非活跃用户的表现；

作者：让更多人看到我生产的内容

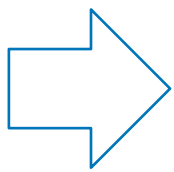
新生产的视频被用户看到的时间越快越好；
在某一时间窗口内被越多人看到越好；
不同等级作者、不同垂直领域作者的表现；

平台：让更多的人来平台看更多内容

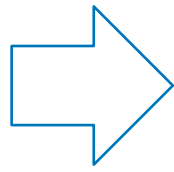
推荐结果准确性、多样性；



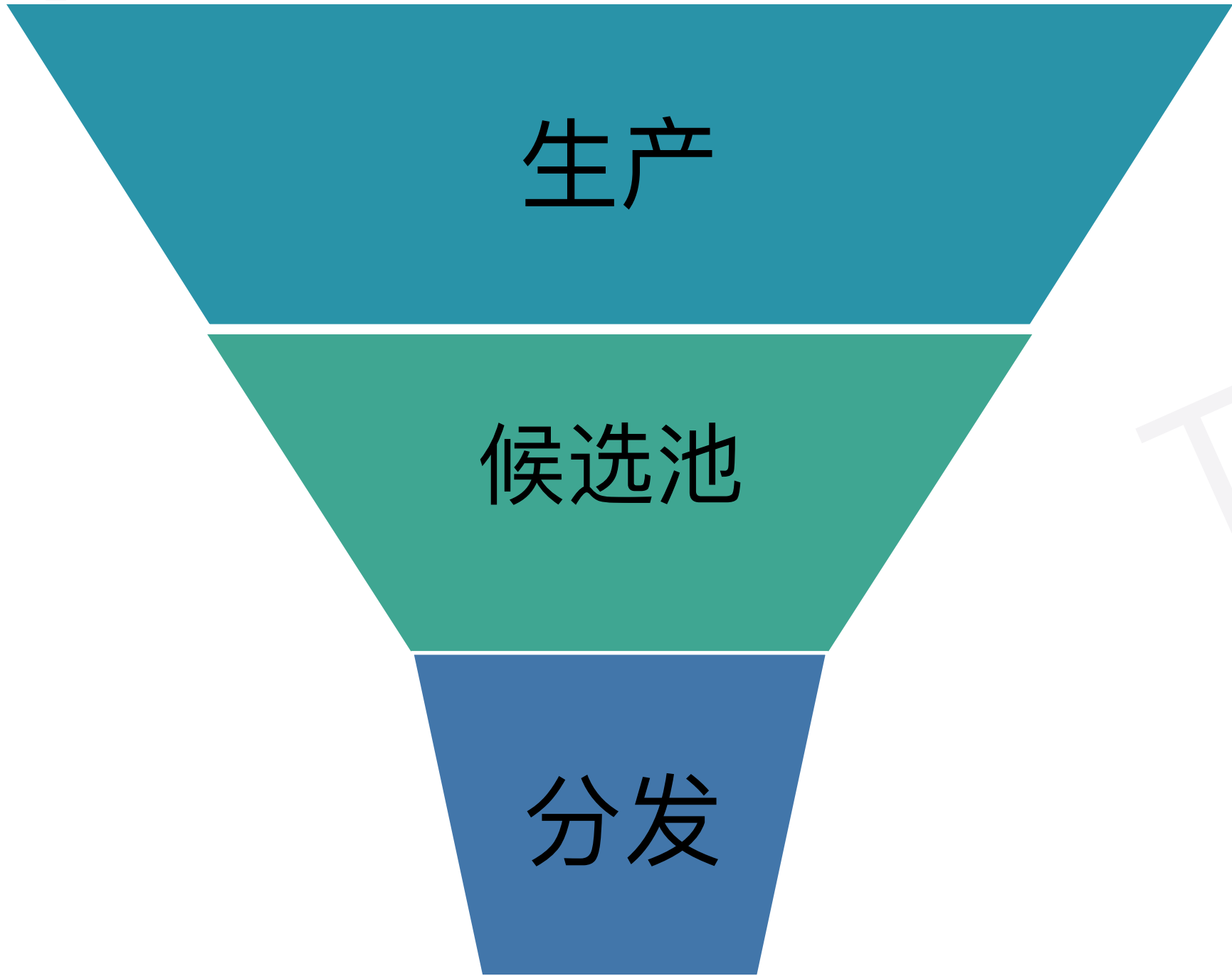
视频生产到分发的漏斗统计



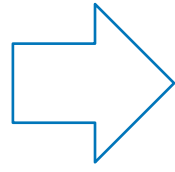
折损原因分析



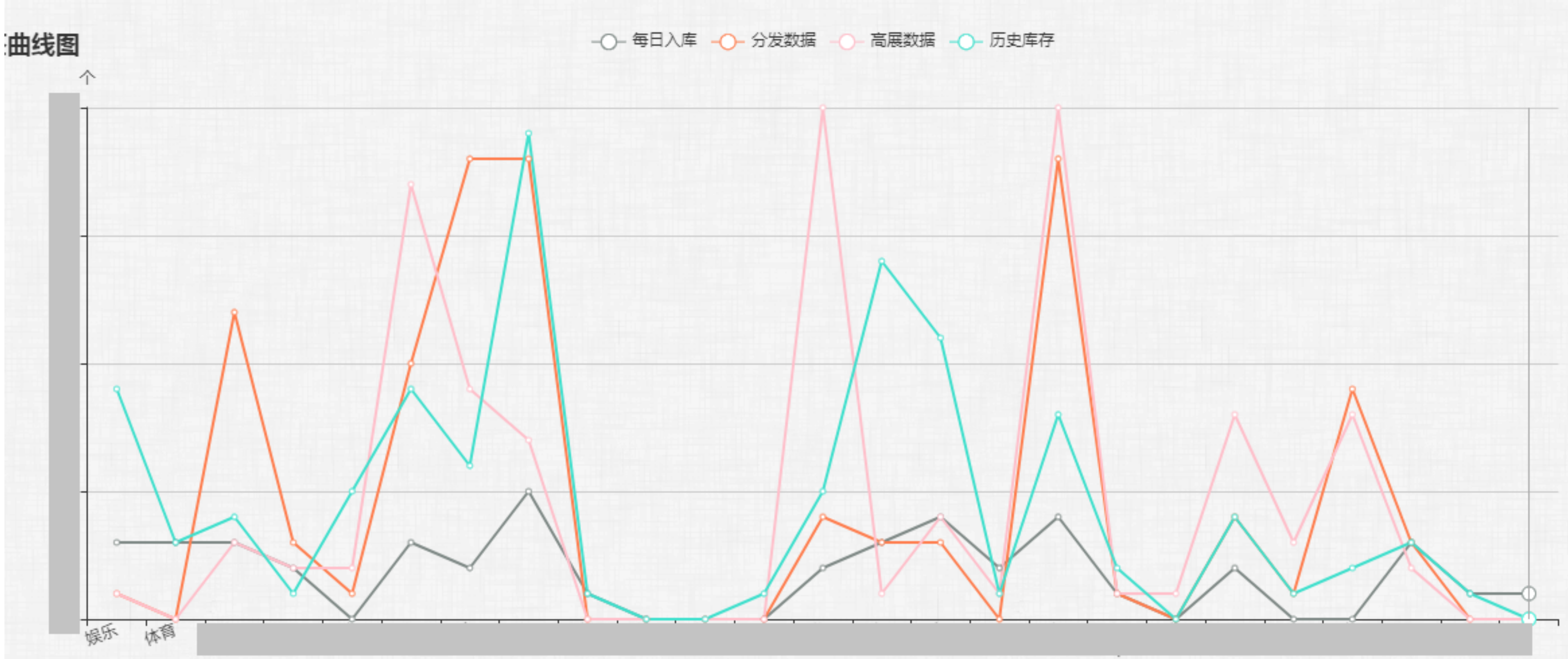
优化解决



标题不规范
封面图不清晰
....



产品设计引导





查询

新增

序号	业务线	区域	名称	用途	功能服务	是否定时	定时规则	结果批次	关联的子任务	提交人	是否发送结果邮件	邮件接收人	新增时间	修改时间	操作
1															结果运行修改
2															结果运行修改
3								次							结果修改
4												@qiya.com			结果运行修改

算法评测平台

评测专项debug分析线下测试线上测试数据定制模板管理报告

评测专项评测任务

创建评测专项

评测专项：输入评测专项名称或id

创建人：输入创建人或修改人

状态：全部

搜索

序号	评测专项id	评测专项名称	创建人	修改人	创建时间	修改时间	状态	操作
1	43				2019-06-05 11:49:49	2019-06-05 11:49:49	草稿	删除 复制 编辑 创建评测任务 评测任务列表
2	41				2019-06-05 10:27:22	2019-06-05 10:27:22	开放使用	删除 复制 编辑 创建评测任务 评测任务列表
3	35						草稿	删除 复制 编辑 创建评测任务 评测任务列表
4	36						草稿	删除 复制 编辑 创建评测任务 评测任务列表
5	34						草稿	删除 复制 编辑 创建评测任务 评测任务列表
6	37				2019-06-03 10:22:05	2019-06-03 10:22:05	草稿	删除 复制 编辑 创建评测任务 评测任务列表
7	38						开放使用	删除 复制 编辑 创建评测任务 评测任务列表
8	32						开放使用	删除 复制 编辑 创建评测任务 评测任务列表
9	33						开放使用	删除 复制 编辑 创建评测任务 评测任务列表
10	29						草稿	删除 复制 编辑 创建评测任务 评测任务列表

上一页

1

下一页

共3页



谢谢

THANKS

