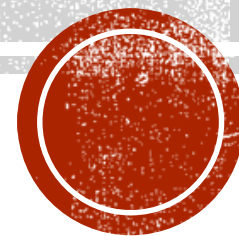


# 易荐 次冷 元启 社区 信息流 推



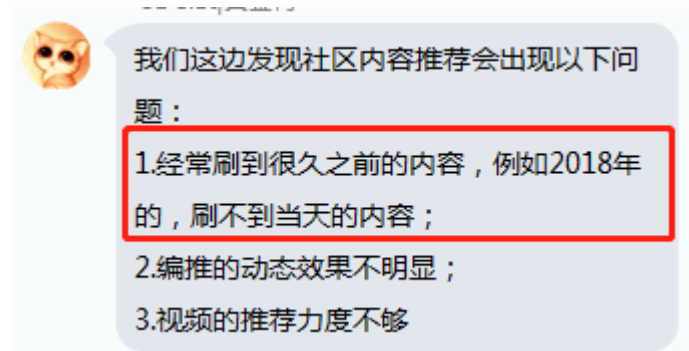
姚昕

战略研究中心 | 数据平台中心 | 数据库营销组

[gzyaoxin@corp.netease.com](mailto:gzyaoxin@corp.netease.com)

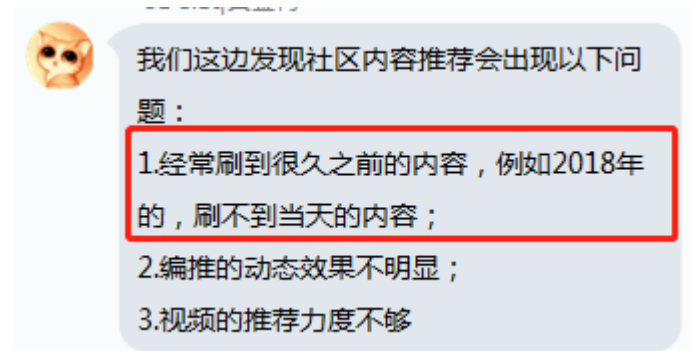
# 目前信息流推荐方案及将来计划

- 第一版现状：
  - 根据用户兴趣召回相关动态，按照日期衰减的热度排序
  - 没有显式表明兴趣的用户，随机召回热门动态
  - 时间紧迫、数据有限，比较简单
- 将来要做的：
  - 协同过滤等
- 存在的问题：
  - 无论是目前的排序还是将来要做的协同过滤的一系列策略，都是基于已有的用户行为记录的
  - 新动态没有任何记录，不管采用任何推荐策略，几乎不会得到任何推荐，永远石沉大海



# 目前信息流推荐方案及将来计划

- 第一版现状：
  - 根据用户兴趣召回相关动态，按照日期衰减的热度排序
  - 没有显式表明兴趣的用户，随机召回热门动态
  - 时间紧迫、数据有限，比较简单
- 将来要做的：
  - 协同过滤等
- 存在的问题：
  - 无论是目前的排序还是将来要做的协同过滤的一系列策略，都是基于已有的用户行为记录的
  - 新动态没有任何记录，不管采用任何推荐策略，几乎不会得到任何推荐，永远石沉大海（推荐的冷启动问题）



# 解决方案

- 流量切分
  - 从现有信息流推荐中切分一部分位置给新增的动态曝光，以获取一定的数据量，让新动态能够根据常规的推荐策略得到推荐
  - 1、常规推荐算法，保证推荐质量
  - 2、新动态试探，保证新内容的曝光度



# 给新动态曝光存在的问题

- 1、UGC(User Generated Content)质量不一，且有些动态不适宜上信息流。若不加区分都尝试曝光给用户，会影响用户体验。
- 2、人工审核不够实时，且费时费力，标准可能不统一。
- 3、动态形式不一（长图文、图片、视频），内容多样，很难用机器自动得到一个准确且度量尺度统一的质量评价。



为何卡黑屏，过不去了？

第二章分4就黑了~~~然后我纠结。为何过不去？

尘沙惑



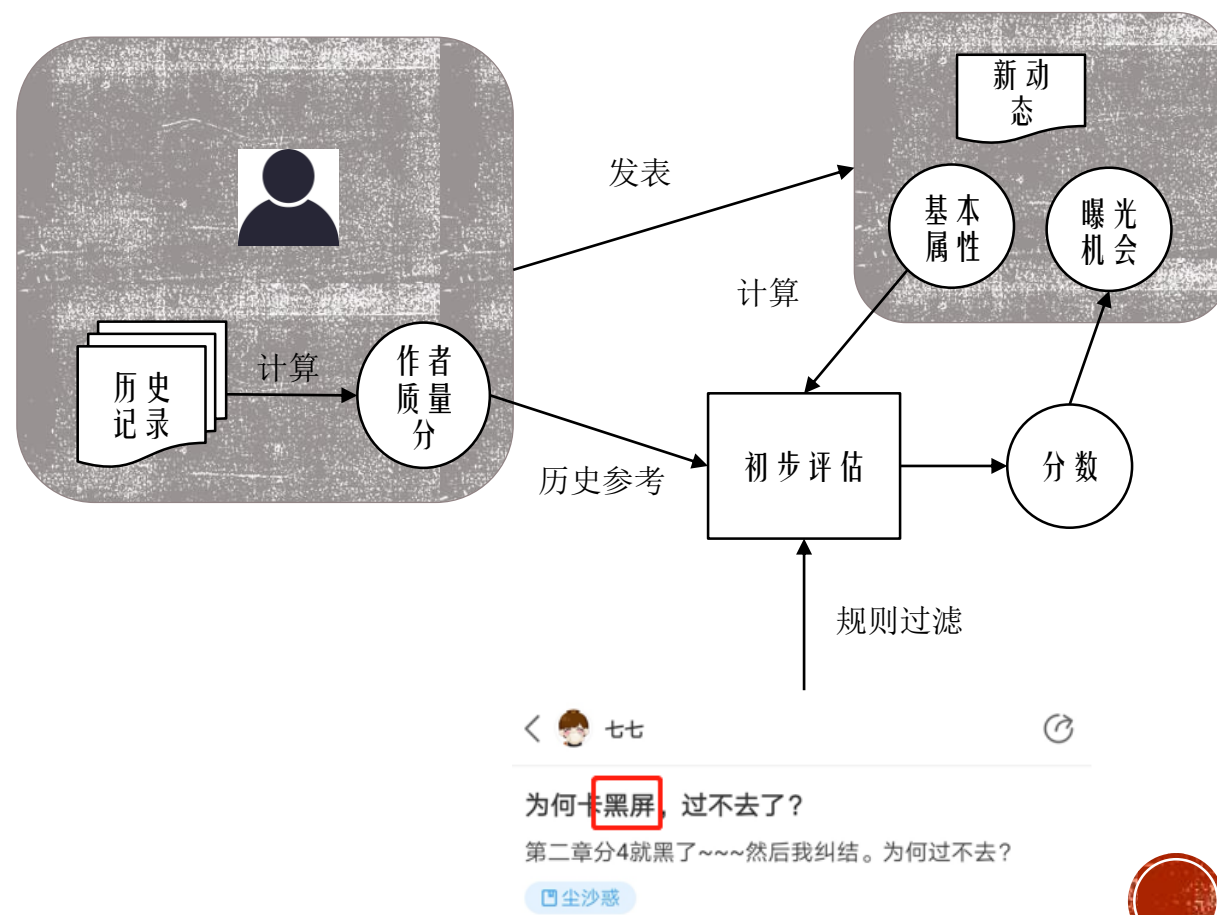
# 应对策略

- 1、简单的机器初步评估
- 2、根据试探性曝光的反馈实时调整
- 3、在前两条基础上的运营适当干预



# 初步评估

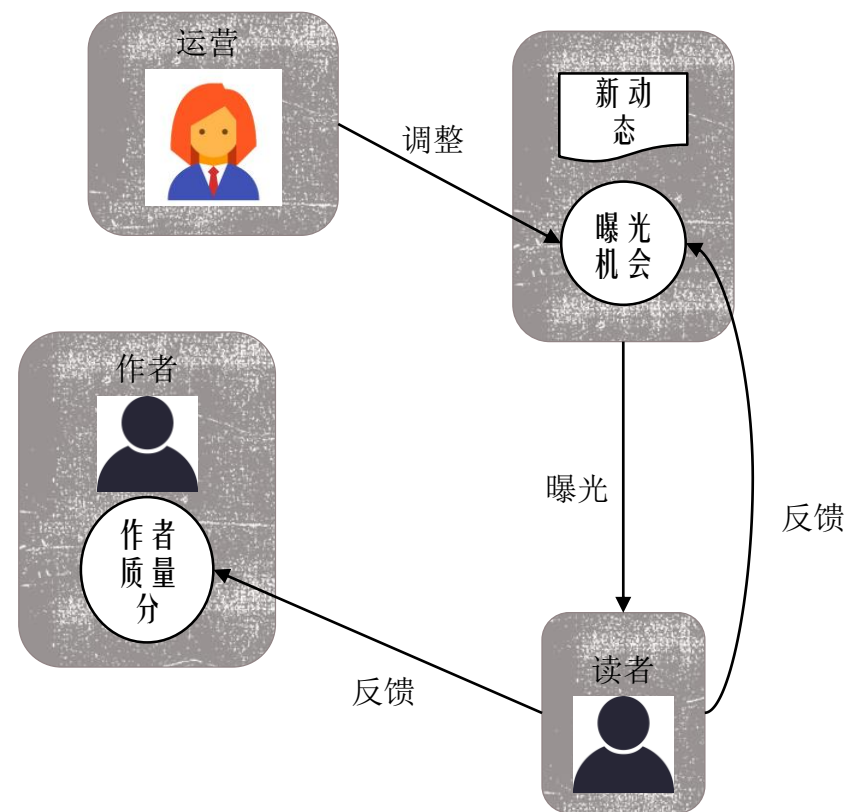
- 用户发表动态时，根据以下信息给出一个初始的分数
  - 当前动态的数据：标题、内容长度、是否带图及数量（长图文、图片）、视频时长（视频）
  - 用户的历史记录：根据过往发表内容的评论率等指标给每个用户计算一个
  - 硬性过滤规则：内容长度低于阈值、包含关键词（黑屏、BUG）等。





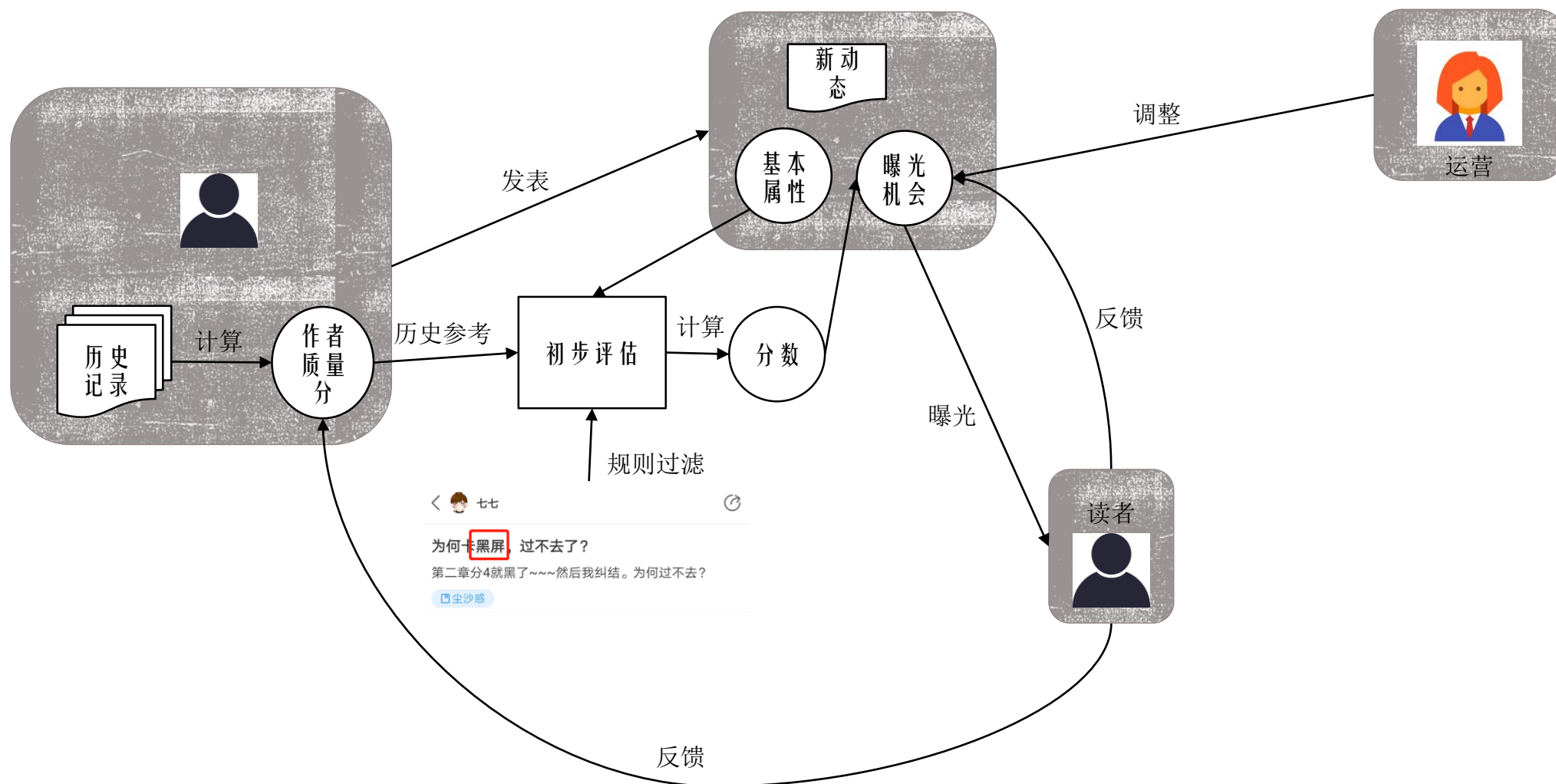
# 动态调整及运营干预

- 根据曝光后的反馈表现来决定该动态后续的曝光机会：
- 新内容会随机获得一些曝光机会（概率）
- 当用户有积极反馈的时候适度增加该动态的后续曝光机会（概率）
- 曝光的结果也会影响作者相应的质量分



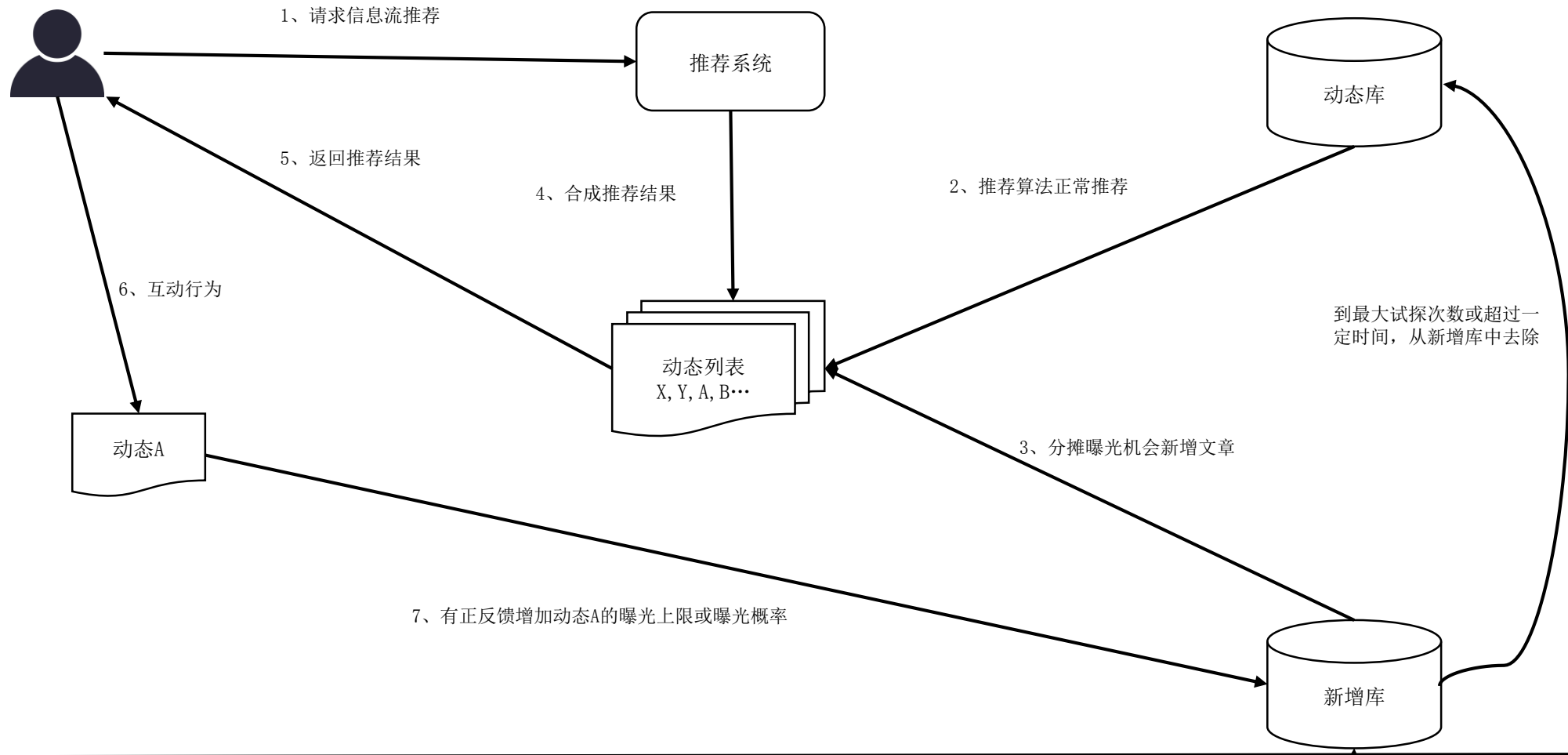


# 策划视角

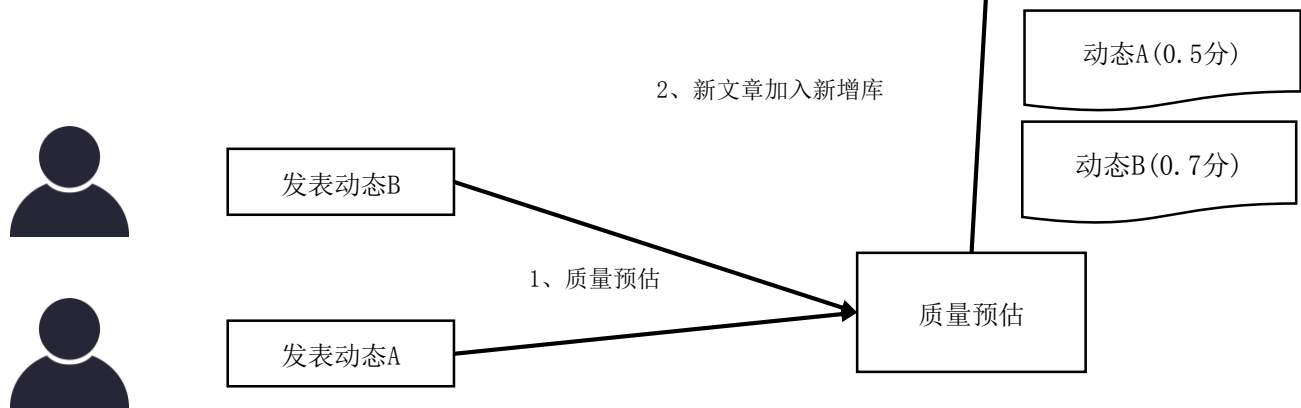


# 程序视角

## 消费过程



## 生产过程



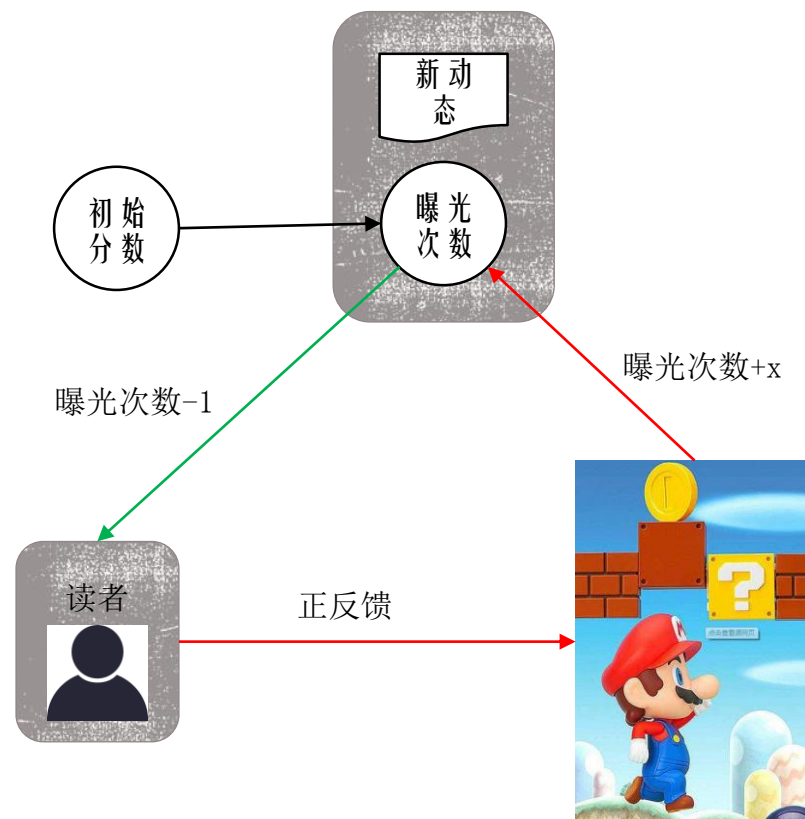
# 一些 问题

- 1、试探性曝光的选择如何量化？
- 2、切分比例与曝光位置的分配
  - 拿出多少比例的曝光量来进行新增动态的？
  - 在页面位置上如何分配？



# 曝光机会如何量化

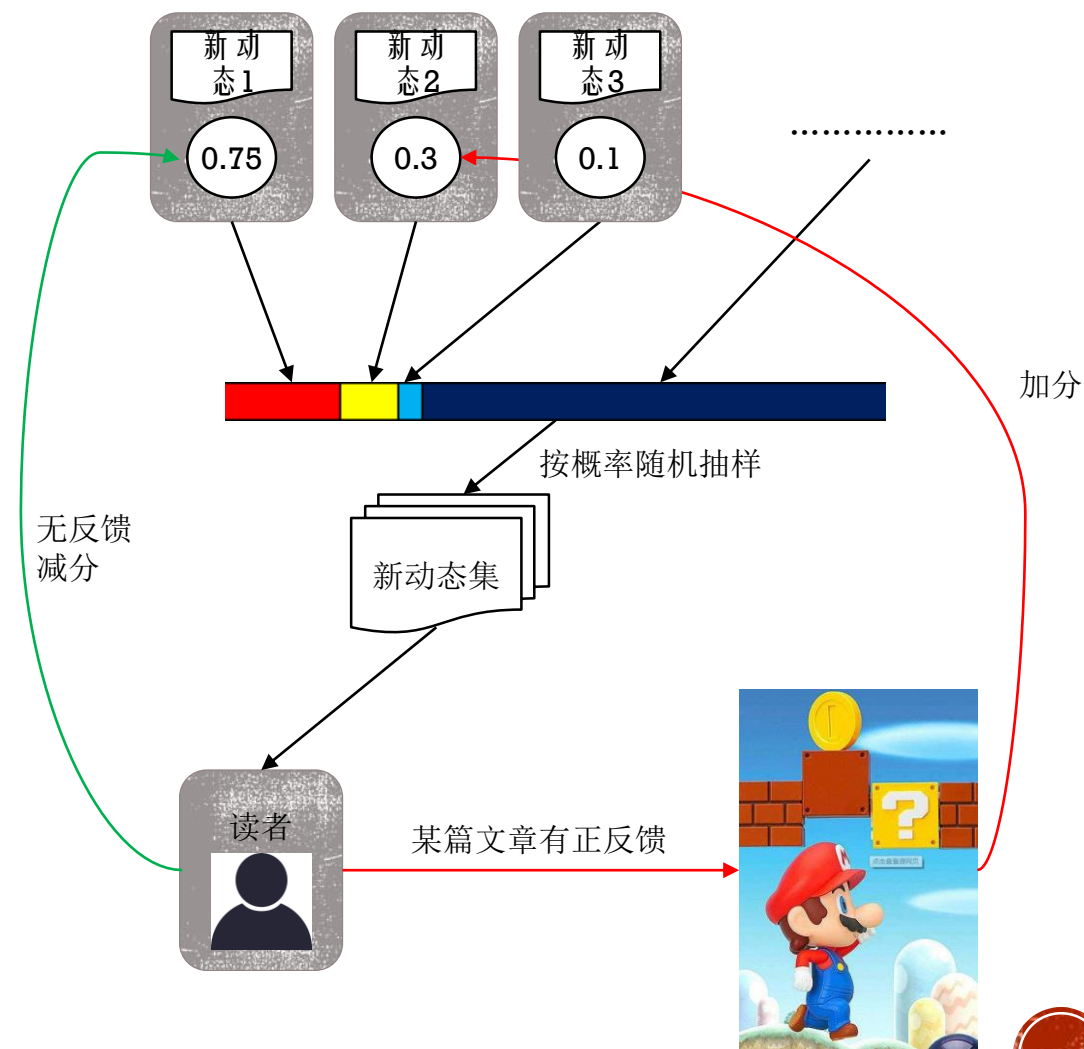
- 一、按曝光次数计算
  - 先根据初始分数计算一个曝光次数
  - 每曝光一次次数减1
  - 每当曝光有正反馈(点赞、评论等)、曝光次数 $+x$
  - 当曝光次数用完或者达到预计的时间，从新增库中去除
- 优点：
  - 数据直观
- 缺点：
  - 参数很难制定
  - 在这个策略之上再考虑用户兴趣（个性化）容易使得策略很复杂
  - 新增文章数及流量密度不规律，影响效果



# 曝光机会如何量化

## ■ 二、按分数计算概率进行抽样

- 曝光概率与文章的分数正相关
- 曝光概率的大小可以加入其它因素的影响
- 文章曝光与否由抽样决定
- 分数根据反馈情况动态的调整
- 到达预定时间，从新增库中去除



# 按分数计算概率进行抽样

- 某篇文章的曝光概率

$$P_i = \frac{s_i * m_i}{\sum_j^N s_j * m_j}$$

- N代表目前新增库的所有文章数量
- $s_i$ 代表文章i的分数
- $m_i$ 代表文章i与当前用户的兴趣匹配程度
- 可以加入其它影响因子

- 优点：

- 可扩展性好，方便与个性化以及其他分配策略混合
- 会根据新增库的大小和流量情况动态的分配

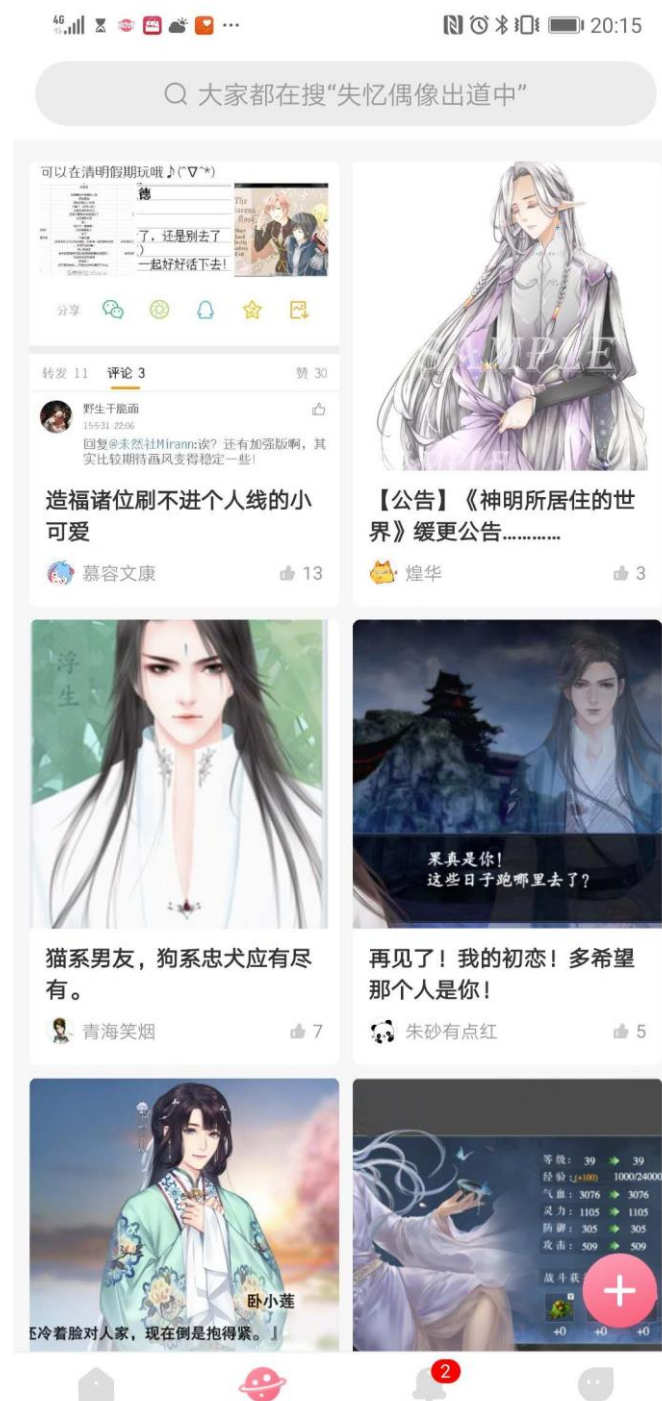
- 缺点：

- 直觉上没有直接指定曝光次数好理解



# 曝光位置分配

- 现状：
  - iOS每次拉取18条推荐
  - Android 每次拉取20条推荐
  - 每屏展示6条动态?
- 切分比例?
- 曝光位分配策略：
  - 固定位置用于新增动态的曝光测试，那么如何分配?
  - 其他策略?





# 其它讨论内容

- 推荐多样化：
  - 动态种类（长图文、图片、视频）多样化
  - 作者多样化？
  - 话题多样化？
  - .....

