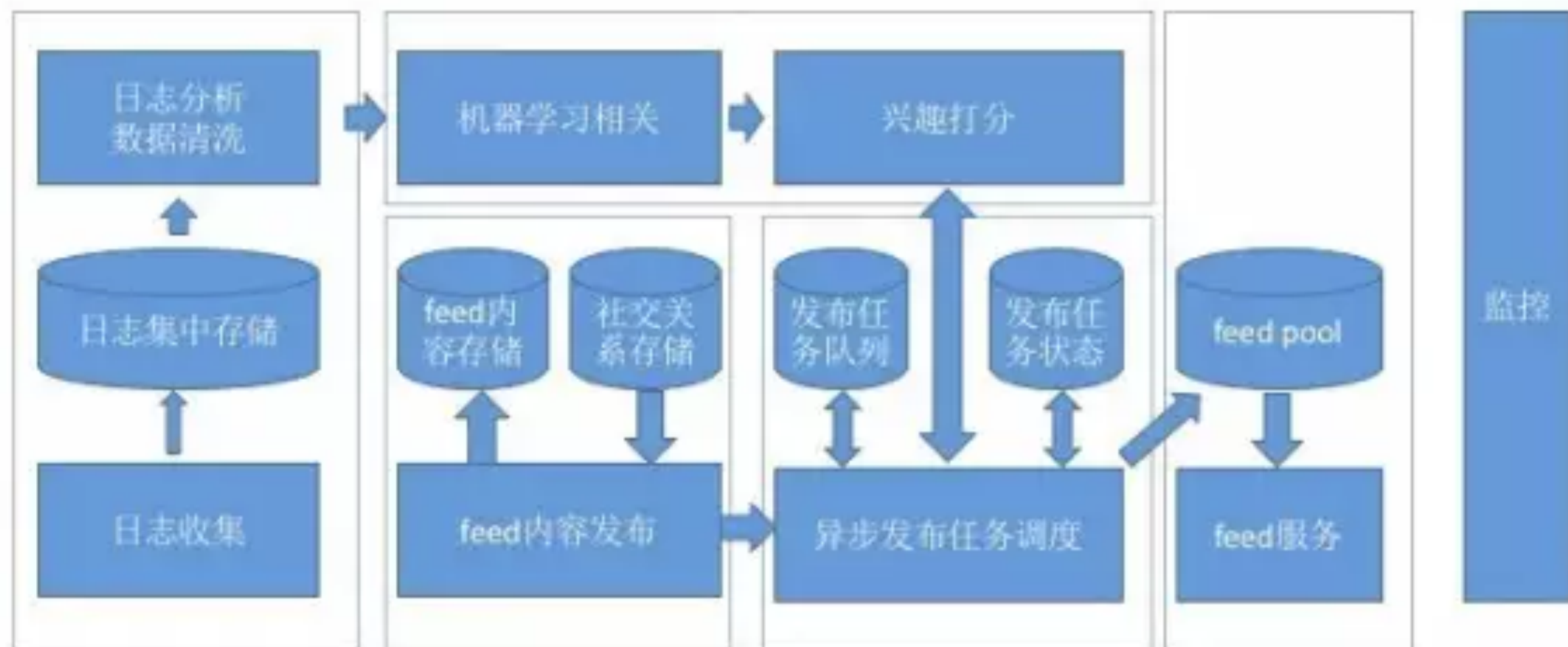
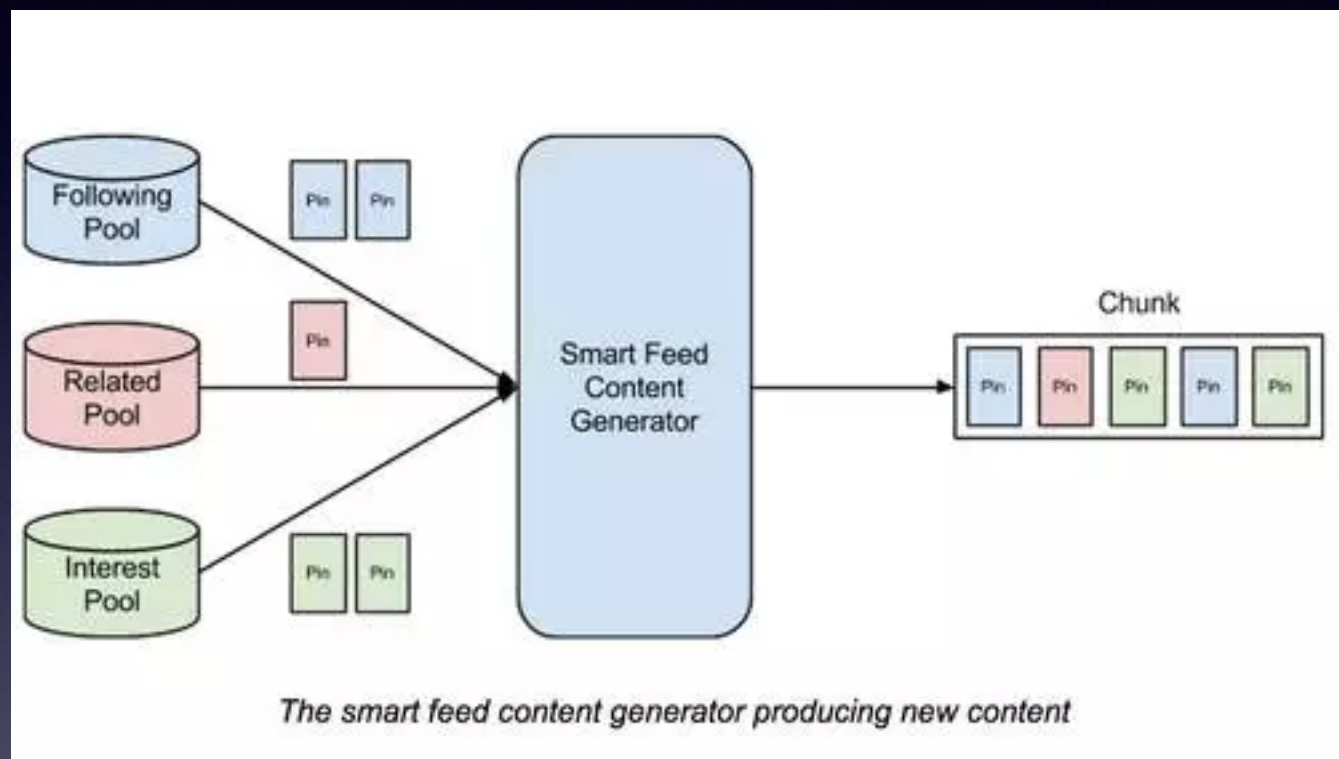
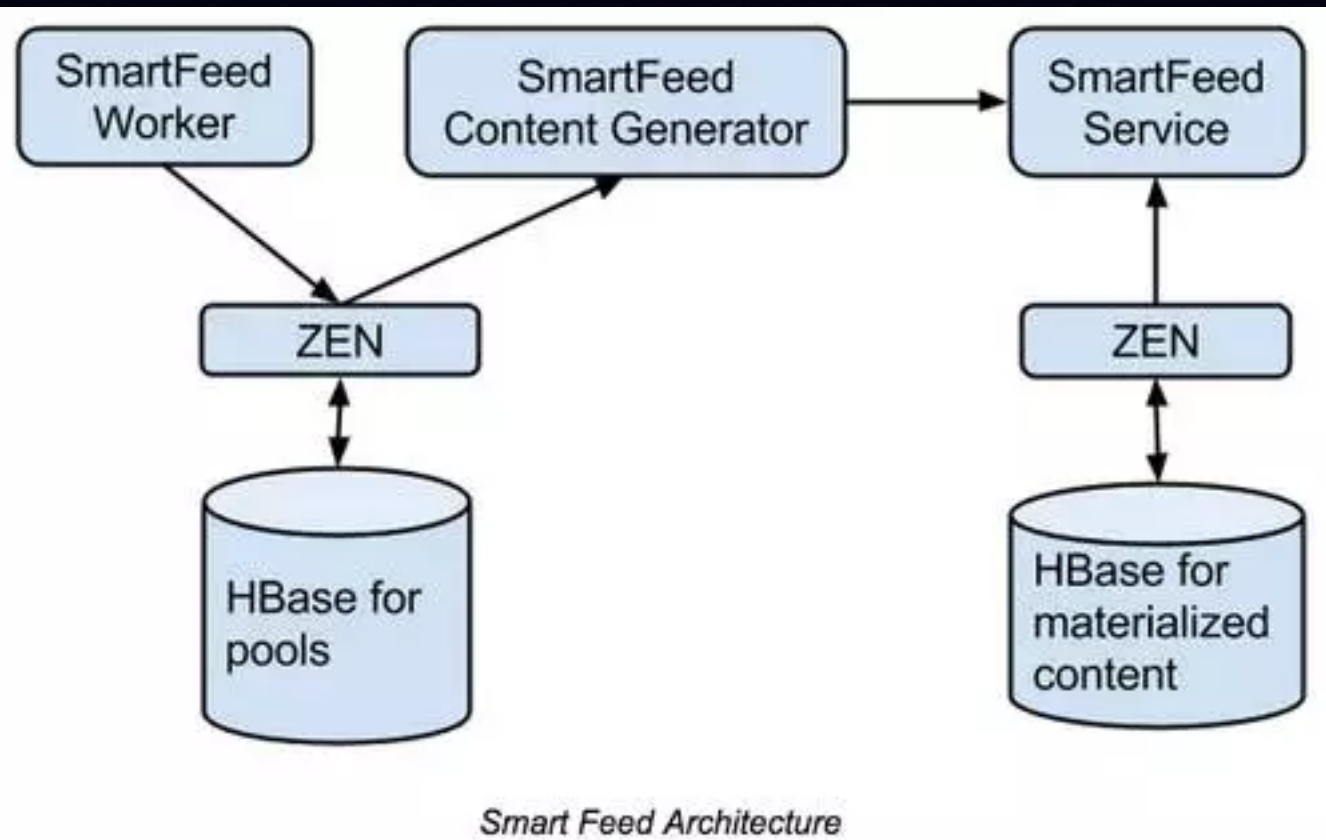


推荐算法

@Tim Qi





缘由

**Facebook日赚千万美元的NewsFeed,
到底是怎么运转的?**

陈开江

- 2013 年之前在新浪微博搜索部和商业产品部任资深算法工程师，先后负责过微博反垃圾、基础数据挖掘、智能客服平台、个性化推荐等产品的后端算法。
- 2012 - 2013 领导翻译了《机器学习：实用案例解析》一书。
- 2013 年末加入传统媒体公司车语传媒，任算法主管，负责从零打造公司转型产品考拉 FM 的个性化推荐系统，如今个性化推荐已成为考拉 FM 与其他 FM 之间最大差异化特性。
- 2015 年初，离职创业，公司拿到 IDG 和晨兴资本的天使投资，产品几经调整，如今专注在用视频分享购物体验，App 名称：边逛边聊。

真相！

早期 Twitter 核心算法

Original Implementation

```
SELECT * FROM tweets
WHERE user_id IN
  (SELECT source_id
   FROM followers
   WHERE destination_id = ?)
ORDER BY created_at DESC
LIMIT 20
```



如何把用户真正最关心的找出来

由拉模式转为推模式

避免两大误区

- 没有目标
- 人工量化

没有目标

- 设计排序算法之前，一定要先弄清楚为什么要对时间序重排？希望达到什么目标？我们这里假设兴趣排序要提高互动率：即希望用户对feed里面的内容互动比例越高越好。所以这里的目标就是：提高互动率。只有先确定目标，才能检验和优化算法。

人工量化

- 常见的产品同学或者运营同学要求对某个因素加权、降权。这样做很不明智，因为**人的知识利于做启发，不利于做量化**。我们很多算法理论上可以得到最优参数，但是所需时间可能是无法承受的，这时候如果有丰富领域知识的人给一些有用的信息，就能大大缩小算法参数的搜索空间，从而在可接受的时间内搜索到效果可以接受的参数。

- 人可以告诉算法很多可能有用的排序因素，缩短效果提升的路径，但是人直接指定参数的权重，对效果提升来说基本上有百害而无一利
- 在算法选定后，人工可以花很多力气在寻找新特征上，这就是传说中的“特征工程”，特征工程还包括对已有的特征进行选择

选择特征的目的

- 控制模型的维度，降低复杂度
- 去掉信号小的特征，提高效果

不能唯数据马首是瞻

不仅仅是“提高点击率”。Facebook 一直是数据驱动的，也是他们能够在争议中把 NewsFeed 坚持下来的信念来源，而不是唯数据马首是瞻，团队内部有很多思考，也有很多变化。

人肉评测小组

非常深入地阐述为什么喜欢 / 不喜欢算法筛选结果，而且会与工程师详细交流评测结果，因为这种人肉评测方式可以有效地拆穿数据说谎，让产品远离一味追求提高数据指标的怪圈。

加强过滤器功能

你可以选择你的动态给谁看，你可以选择不看谁的动态（Hide），你也可以选择优先看到谁的动态（See first），当然你可以取消关注一些人或者公共主页，将这些控制权交给用户，一方面安抚了用户被算法接管的不安，另一方面也是一项重要的数据来源。事实上，如今的过滤器已经复杂到大多数用户都不会使用这些控制权。

如何衡量用户真的感兴趣

人的非理性占据了绝大多数时候，而“不一致性”又是非理性的最重要表现，根据心情不同，用户有不同的感兴趣的标准，而且这些标准也许还是相互矛盾的。

两大方面

- 社交，腾讯
- 算法，今日头条

微信商业化

- 把底栏四个功能入口中的“通讯录”拿掉，换成“动态”，整合朋友圈和公众号内容。
- 把 Facebook 已经趟过路的四种广告形态 Suggested Page、Suggested Post、Suggested App、以及 Video Ads 搬过来。
- 广告算法团队把流量变现大神张栋找过去。
- 齐活，等着数钱！News Feed 广告变现最相关的是使用时长，就算我们不考虑微信变态高的用户使用时长数据，我们相比 Facebook 随便给打个对折吧，保守估计此块业务的营收每日 2500 万美金是完全可以期待的！

My Point This Time

把选择权交给用户

不对用户做评价

人在江湖，身不由己

避免

不喜欢就撤下去了，再给你换别的菜吃，不满意还可以再换，直到你同意为止。

轮子

Stream-Framework

Stream Framework is a python library which allows you to build activity streams & newsfeeds using Cassandra and/or Redis.

<https://github.com/tschellenbach/Stream-Framework>

Examples of what you can:

- Activity streams such as seen on Github
- A Twitter style newsfeed
- A feed like Instagram/ Pinterest
- Facebook style newsfeeds
- A notification system

当然没必要使用这个，因为用了也没什么效果