

# 电商用户的兴趣模型

赵海臣

# 背景

🌀 做千人千面也有将近一个年头，这一年中迭代了各种各样的算法

- ★ 基于统计的标签推荐
- ★ 基于als mf的长期行为个性化推荐
- ★ 基于item-based cf的短期行为个性化推荐
- ★ 用户标签(品牌、分类)个性化推荐
- ★ 基于订单关联规则的推荐
- ★ 用户长期兴趣域推荐
- ★ 新用户冷启动推荐
- ★ 线上实时推荐

🌀 经过这些算法迭代，对用户的兴趣模型有了比较清晰的认识，大致可以总结出电商用户的兴趣模型。

# 用户基础兴趣偏好

🌀 用户基础兴趣主要是基于用户的固有标签的统计学兴趣，例如不同地域、不同贫富程度的人有着不同的兴趣偏好

★ 不同地域的兴趣偏好：

- 对于化妆品来说，干燥的北京与湿润的成都，两个地方的用户对护肤品的需求不一样

★ 不同贫富程度的兴趣偏好：

- 对于化妆品来说，富裕的人倾向于购买较为高端的品牌

🌀 新用户推荐主要基于这两个可以获取的信息——手机平台(Android, iphone, 根据统计, iphone用户购物习惯与Android用户有明显差别)、从ip解析出地域信息(不同地域的用户的购物偏好也有区别)

# 用户基础兴趣偏好推荐

✎ 用户基础兴趣偏好主要用于推荐没有任何历史交互信息的新用户，主要基于用户画像统计学的推荐

✎ 用户基础兴趣偏好推荐的优缺点

- ★ 优点：用户基础兴趣偏好能够拟合用户的模糊兴趣偏好，能够以最少的信息，获得一定的收益。往往用于新用户推荐中，ctr、召回率、精确度、f值像对于没有任何推荐来说，有可观测到的一定提升，甚至超过了运营的选品推荐。
- ★ 缺点：用户基础兴趣偏好推荐的精确度非常低，它只是一种兴趣域的模糊轮廓，ctr、召回率、精确度、f值是所有推荐方法中最低的。

# 长期兴趣域

每个用户都有一个基础偏好，这反映在用户的长期(数月数年)购物偏好上，例如

- ★ 男性跟女性的长期兴趣差别
  - 男性偏好男性护肤品
  - 女性偏好女性护肤品
- ★ 不同职业的长期兴趣差别
  - 工程师偏好于更实用的功能性商品
  - 艺术家偏好于更浪漫的视觉性商品
- ★ etc

在长期模型上，我们需要侧重于挖掘出用户的泛化兴趣域，而非精准的单个商品

- ★ 使用基于用户-商品交互行为的聚类技术，将用户的心理兴趣域给聚类挖掘出来
- ★ 基于用户的心理兴趣域对用户进行心理兴趣域的刻画排序，例如使用ALS MF推荐方式对用户的心理兴趣域进行刻画排序。

# 长期兴趣域推荐

✎ 获得了用户的长期兴趣域后，再基于用户的长期兴趣域对商品进行推荐，详情参考“用户长期兴趣域推荐模型”

✎ 可以使用商品的品牌、分类客观兴趣域代替用户的主观兴趣域，即标签(品牌、分类)推荐

✎ 长期兴趣域推荐的优缺点：

- ★ 优点：长期兴趣域推荐捕获的是用户的长期的兴趣偏好，具有稳定、较大泛化域的优点，能够推荐出大范围的用户可能感兴趣的物品，ctr、召回率、精确度、f值相对于无推荐的物品有明显的增长，甚至能超过热销榜。
- ★ 缺点：由于长期兴趣域推荐捕获的是大泛化兴趣域，因此，推荐的精确度不高，ctr、召回率、精确度、f值相对于其它精准推荐效果要差。

# 短期关注焦点

☞ 用户的短期行为代表了用户近期的关注焦点，这种焦点更多地是临时、且较为急迫的需求，这主要反映在用户最近一段时间(数天到数月)的行为上，例如

- ★ 计划旅行，在正式旅行的前一段时间会关注旅行相关的商品
- ★ 考试临近，会关注一些考试相关的书籍或者用品
- ★ 计划添置某种电器，会关注相关的电器类商品
- ★ etc

☞ 在短期行为推荐上，我们需要侧重于挖掘用户的细致商品需求，猜测用户的感兴趣的商品

- ★ 使用基于cf的短期模型，对用户的感兴趣的商品进行挖掘、推荐
- ★ 使用基于用户近期点击的模型，捕捉用户的核心兴趣商品
- ★ 使用基于用户订单的关联规则模型，捕捉用户的短期内关联性需求。



# 短期关注焦点推荐

短期关注焦点推荐主要应用融合点击模型的item-cf作为推荐算法，能够较为快速地捕捉用户的短期关注焦点，详细参考“聚美cf协同过滤”

与此同时，用户短期内购买过的商品的补充性需求也是用户短期内的关注点，使用关联规则算法捕捉用户的短期购买补充性需求，详细参考“基于订单关联规则的推荐”

短期关注焦点推荐优缺点：

- 优点：短期关注焦点推荐能快速捕捉用户当前的关注点，作为用户关注焦点，其推荐的商品有较高的精准性，ctr、召回率、精确度、f值皆远超长期模型。
- 缺点：短期关注焦点推荐容易造成马太效应，因为短期模型推荐出的商品有较高的点击率，因此导致短期模型对用户的兴趣更进一步收敛到焦点商品上，容易导致用户的疲乏
  - 例如，uc头条的例子，一些很吸引眼球的野史很容易让人点进去，后来慢慢全部是野史，用户欲罢不能地继续点击，但其实用户对这些内容已经很反感。这些内容能够获得较高ctr，但是用户整体点击次数却下降了。
  - 又如知乎上，老是推送近期的焦点(spark)，忽视了长期的个性轮廓(科技、军事)，容易导致用户反感。



# 实时需求

🌀 用户的实时浏览、搜索、加购、订单行为反映了用户当前的最迫切的需求，是距离用户需求最近的一个行为表现，抓住了用户的实时需求，能够最大化ctr、召回率、精确度

🌀 用户的实时行为的重要性不言而喻，是用户当下的实际需求与兴趣反映，用户会主动浏览、搜索、购买自己当下最想要的东西，无论是长期模型、短期模型都是在模糊地猜测用户的当前需求，而用户的实时行为就是这个当前需求的最直观体现

- ★ 用户需要香皂，就会搜索香皂相关的
- ★ 用户需要面膜，就会搜索面膜相关的
- ★ 用户需要洗发液，就会搜索洗发液相关的
- ★ 当然，用户也可能被广告式启发地对某些看到的商品感兴趣，但大部分用户的购物行为相对来说是有目标性的。

# 实时关注点推荐

☞ 用户的实时关注点体现在用户的实时浏览、搜索、加购、订单等行为上，抓住了这些实时点，往往能达到长期模型、短期模型都无法比拟的高ctr、召回率、精确度

## ★ 基于搜索词的推荐

- 用户搜索完一个关键词后，我们通过统计历史上该关键词搜索完用户购买商品的频次，按频次来进行优先级推荐。
  - 搜索词的推荐有效期40分钟

## ★ 基于购物车/订单的推荐

- 用户将某个商品加入购物车，或者生成订单后，通过关联规则、订单cf来产生补充性的商品推荐。
  - 购物车推荐有效期10分钟
  - 订单推荐有效期18分钟

# 实时关注点推荐

🌀 实时关注点推荐的优缺点：

- ★ 优点：实时关注点推荐能够最精确地捕捉到用户的当前需求，具有长期模型、短期模型都无法比拟的高精确性，因此实时关注点模型的ctr、召回率、精确度都高于长期模型、短期模型
- ★ 缺点：实时推荐对机器性能要求高，并且用户的实时需求并不一定反映用户的个性兴趣偏好，很多时候是用户的临时性需求，时间延续预测性最差。

# 电商用户的兴趣模型结论

兴趣模型如果用一个二维兴趣图(类似于黑洞二维表示那种)来表示：

- ★ 用户的基础兴趣偏好是大图上隐约可见的平缓凹凸；
- ★ 用户的长期兴趣域模型是大图上明显可见的次平缓凹凸；
- ★ 用户的短期兴趣模型是大图上明显的较急促的凹凸；
- ★ 用户的实时关注点模型是大图上的针尖状凹凸

从时域上看

- ★ 用户的基础兴趣偏好—隐约可见的平缓凹凸—基本上稳定不变
- ★ 用户的长期兴趣域模型—明显可见的次平缓凹凸—变化缓慢
- ★ 用户的短期兴趣模型—明显的较急促的凹凸—快速变化
- ★ 用户的实时关注点模型—针尖状凹凸—转瞬即逝

**THE END**

**THANK YOU!**