

千人千面排序

赵海臣

ECPM

- ▶ eCPM (effective cost per mille) 指的就是每一千次展示可以获得的广告收入，展示的单位可以是网页，广告单元，甚至是单个广告（在 AdSense “高级报告” 的 “数据展示依据” 下拉框中可以选择）。
 - 默认情况下，eCPM 指的都是千次网页展示（Pageview）收入。eCPM 只是用来反映网站盈利能力的参数，不代表收入。
 - 销量为M，曝光次数为y，则ECPM为：

$$ECPM = 1000 * \frac{M}{y}$$

搜索所用ECPM

- ▶ 某词搜索下销售量为 M ，某次搜索销售量曝光为 y ，ECPM(传统ECPM*10)为：

$$ECPM = 1000 * \frac{M}{y} * 10$$

- ▶ 基于置信度的调整，乘以置信度基准，若曝光量大到一定程度($y^{0.75} = 1000$)时，不对ECPM进行衰减，取1，否则对ECPM根据低置信度进行衰减：

$$ECPM_{4Search} = ECPM * \min\left\{1, \frac{2 * y^{0.75}}{1000 + y^{0.75}}\right\}$$

标准分数

- ▶ 标准分数（standard score）也叫z分数（z-score），是一个分数与平均数的差再除以标准差的过程。用公式表示为：
 - $z = (x - \mu) / \sigma$ 。其中x为某一具体分数，
 - μ 为平均数， σ 为标准差。
- ▶ Z值的量代表着原始分数和母体平均值之间的距离，是以标准差为单位计算。
 - 在原始分数低于平均值时Z则为负数，反之则为正数。

使用标准分数处理ECPM

- ▶ 限制z值取 $[-3, 3]$ 区间，超出部分进行截断，并取+5偏置，使得标准分数ECPM取值在 $[2, 8]$ ：

$$STDSCORE = 5 + \min \left\{ \max \left\{ \frac{ECPM4SEARCH - \overline{ECPM4SEARCH}}{\delta}, -3 \right\}, 3 \right\}$$

排序分数

- ▶ 由“ECPM+个性化分数+曝光量降权”三部分构成：

$$RANK = a * \frac{Time - 10}{1 + (Time - 10)} * STDSCORE + b * PersonalizationScore + c * e^{-\frac{y}{100}}$$

- 因为以上午10点作为基准，因此STDSCORE(ECPM)的置信度权重随着10点逐渐上升；
- Personalization分两种：
 - 计算商品个性化分数时：ALS输出的user-item个性化分数
 - 计算专场个性化分数时：ALS输出的user-brandId或user-catg_v3_3标签个性化分数计算专场分数(详见下页PPT)
- 曝光量降权：
 - y为曝光量，当一个商品曝光量大时，为了挖掘长尾，对曝光量大的商品进行适当降权。
- a,b,c为三个部分的调节参数。

个性化专场分数的计算

- ▶ 一个专场由很多不同catg与不同brand的商品构成, 计算一个专场的整体PersonalizationScore需要结合专场内的所有商品:

$$PersonalizationScore = \sqrt{\sum TagPersonalScore^2 * \frac{TagNum}{TotalNum}}$$

- TagNum是指专场内对应标签的商品数量
- TotalNum是指专场内所有商品的数量

强制占位

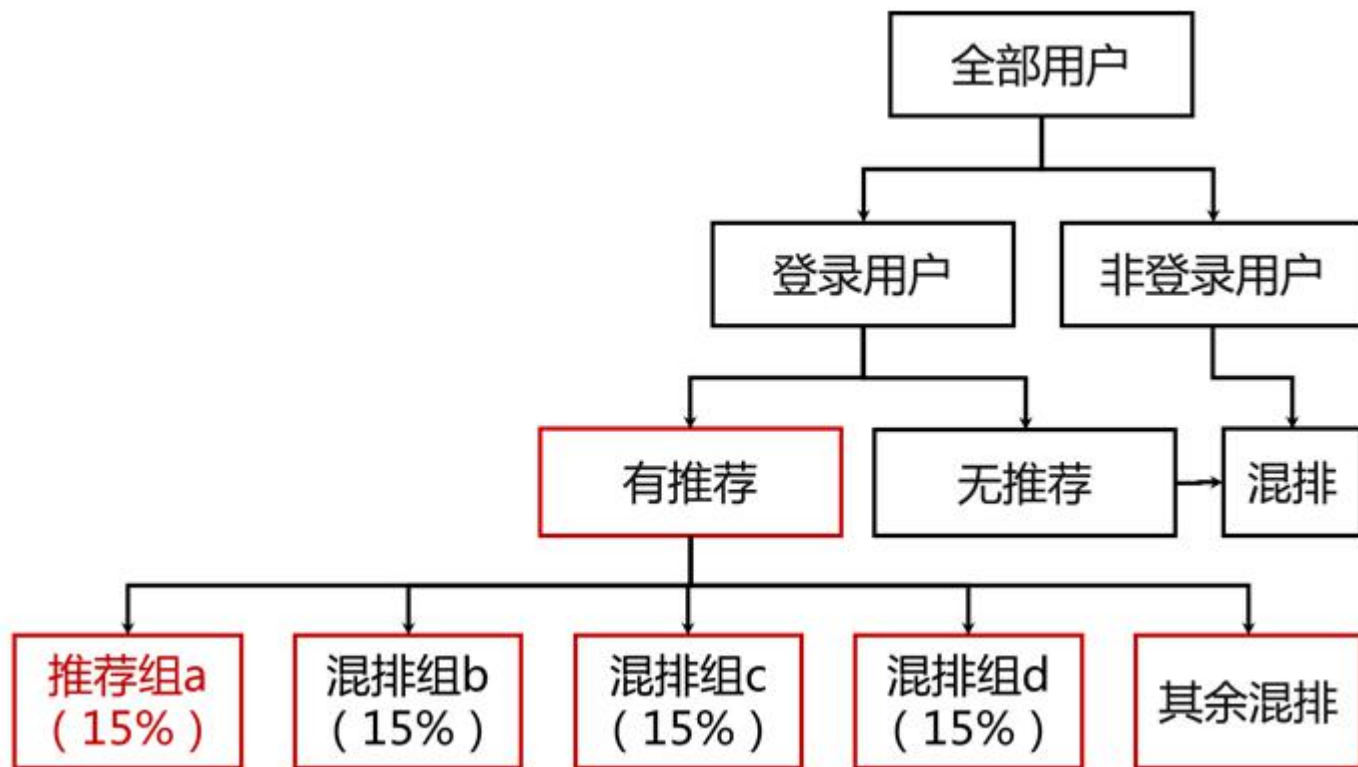
- ▶ CF输出的短期商品，暂时取各个用户前3名强制占据初始列表的2，4，6号位，其后续排序分数按照前述公式进行。若ALS无对应分数，则其个性化分数为0。

在线测试

▶ 实验测试方案

◦ 流量分配方案

- 登录且有推荐用户的15%作为实验组，设立三组登录且有推荐的15%用户组委混排对照组。



浏览分配结果

- ▶ 实验组相比对照组提升比例：
 - 取三组对照组的平均值作为对照组基准

数据类别	销售额增长	销售量增长
Android	56.1%	44.0%
iOS	73.9%	80.8%
综合	67.0%	63.1%

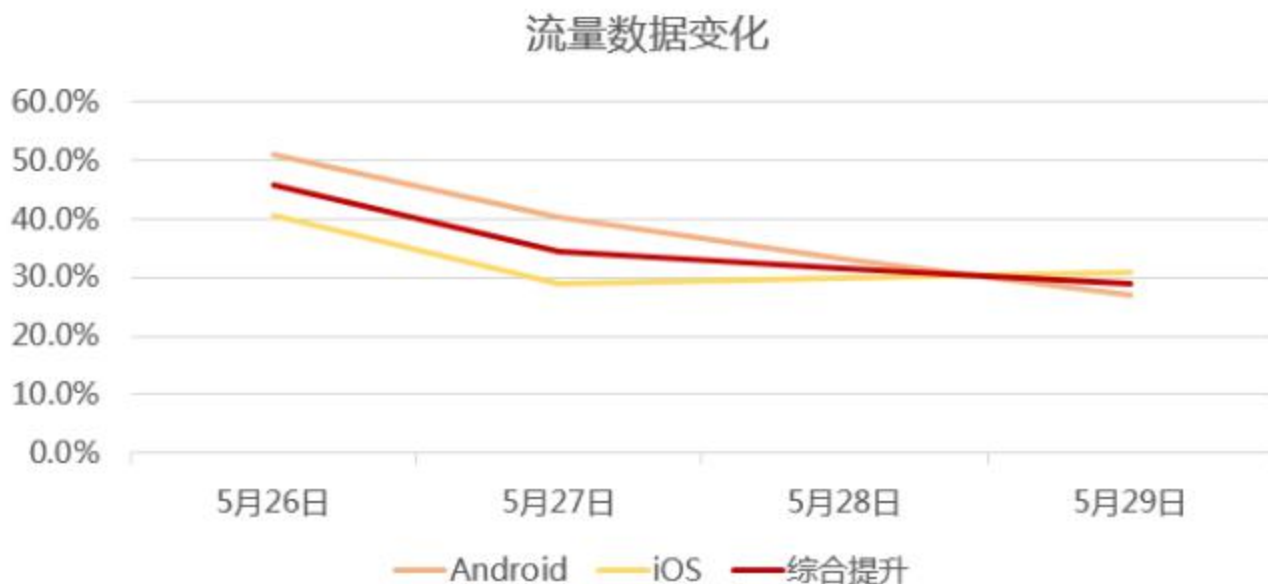
浏览数据分析

- 曝光次数 / UV的提升，用户因为千人千面，用户浏览了更多的素材。
- 点击人数 / UV的提升，代表因为千人千面，更多用户产生了点击行为
- 点击次数 / UV的提升，代表因为千人千面，平均每个用户点击了更多的素材

数据说明	曝光次数/UV	点击人数/UV	点击次数/UV
Android	17.2%	13.9%	38.0%
iOS	12.4%	8.0%	33.5%
综合提升	14.7%	11.1%	35.8%

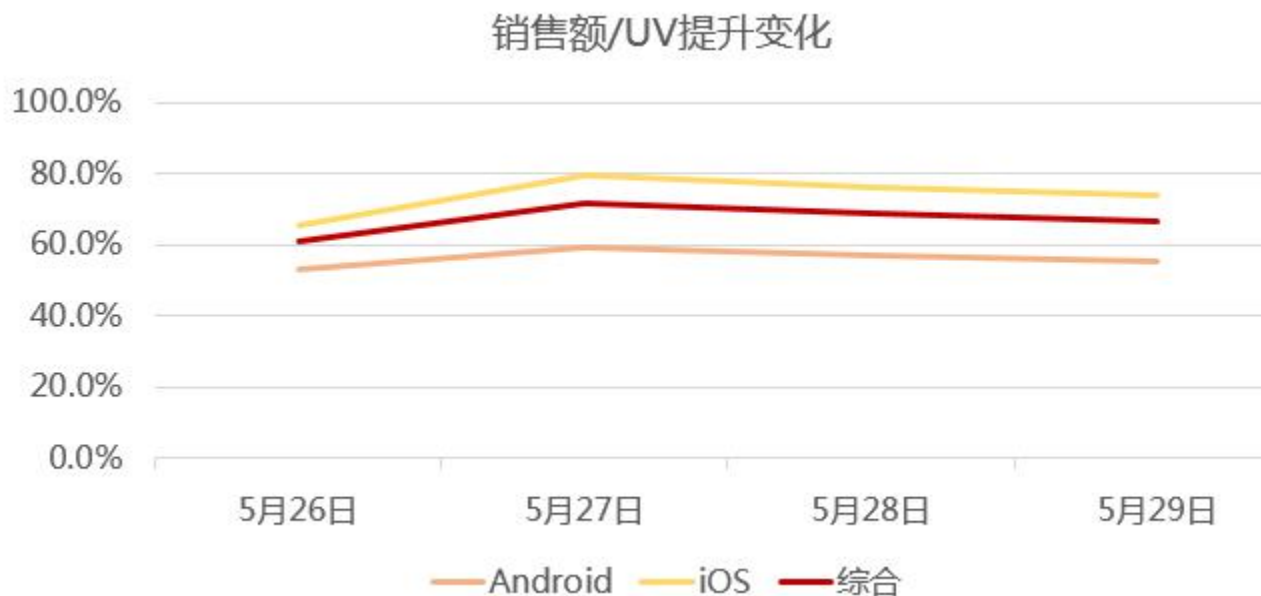
分天数据提升

- ▶ 流量数据为点击数/UV，基本稳定提升30%以上，除了第一天上线数据流量数据偏高，其余数据没有明显衰减。



分天数据提升

- ▶ 销售数据取销售额/UV查看变化趋势，基本稳定提升60%以上，除了第一天上线数据流量数据偏低，其余数据没有明显数据波动。



The end