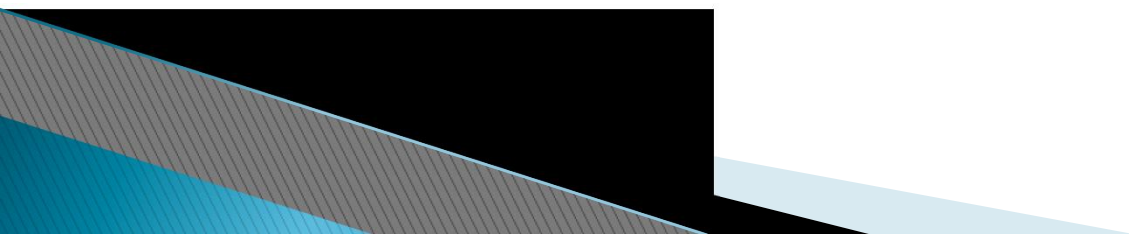


ALS参数调试

赵海臣

前提背景

- ▶ ALS运行一次需要大量的时间，而公司的集群资源有限，需要以最少的次数尝试出较优的结果，而不能用随机组合方式。



ALS待调参数

参数	最优值
浏览数据时长	90天
订单数据时长	180天
购物车数据时长	90天
心愿单时长	90天
复购过滤时长	15天
召回商品数量	5000个
浏览数据权重	1
订单数据权重	2
购物车数据权重	2
心愿单数据权重	2

ALS待调参数

参数	最优值
隐含行为参数 α	10.0
Als维度 rank	200
Als迭代次数 iteration	20
Als校正因子 λ	1.0
时间衰减参数 ω	0.01

目标评判公式

- ▶ 以历史某日作为推荐标准日，如2017-03-15，以2017-03-15的信息为标准生成推荐列表。
- ▶ 从推荐列表中抽取100,000用户来预测接下来一个星期的用户真实购买行为，包含：
 - 召回率：
 - $\text{用户真实购买了的预测商品数} / \text{用户真实购买的商品总数}$
 - 精确度：
 - $\text{用户真实购买了的预测商品数} / \text{用户的预测商品总数}$
 - 销售额：
 - 用户真实购买了的预测商品的总价

参数调试原则

- ▶ 参数初始化：
 - 根据经验给各个参数一个经验值
- ▶ 参数调试规则：
 - 首先对各个参数分别进行粗调，每个调试参数10倍的差距，例如AIs校正因子 λ 给予[0.01, 0.1, 1.0, 10.0]，找出最佳参数的粗略点
 - 再对各个参数进行细调，发现 λ 在1.0取得最优值，再分别调试[0.5, 1.5]，发现1.0仍是最优值，可以继续调试[0.8, 1.2]
- ▶ 根据参数调试原则分别调试各个参数，能保证结果一直在往最优参数收敛。

参数优值判定标准

- ▶ 以召回率、精确度、销售额评判结果高的为较优参数
 - 实际调试过程中发现，召回率、精确度、销售额呈现共同的波动方向，取销售额作为最终评判标准，因销售额信息中既包含销量、也有利润的因素。
- ▶ *单参数调试的缺点：
 - 没有考虑参数的共同作用，例如有些参数是与其它参数协同发挥作用的，单次调试一个参数无法获得共同最优值。
 - 改进方法：使用多参数共同随机组合调试，但是需要的运算次数、调试成本将大大增加。

最大影响参数

▶ 1. 召回商品数量:

召回商品数量	预测判定销售额
1000	3993373.653138
3000	4513844.423509
5000	4087836.383076
7000	3320126.122041

▶ 2. als维度rank:

- Rank 越大，预测判定额越大，但受限于集群性能，尚未达到极值。

最大影响参数

- ▶ 3. 复购限制时长：
 - 复购限制时长将很大程度降低预测的销售额，原因很可能是用户复购情况很常见。

基于回归分析的参数调试

- ▶ 将所有的调试记录使用线性回归回归出线性方程，再从线性方程上找极值。

The end