

# 点击率预估方案

---

- 线下评测方案

方案一：取测试集前一天的数据并数据均衡化后作为测试集； 方案二： xfold

- 特征工程

假设官方给出训练集给出的数据时间范围是train\_start、train\_end，  
取train\_start、train\_end之间的train\_middle1和train\_middle2作为模型训练集，令num\_day = train\_middle1-train\_start。

特征按照重要性分为1、2、3级，其中3级是最直观的特征，必须加入；2级为属性特征，建议加入；1级可能会有作用，可以尝试加入。

- 基础信息(2级) 用户id

性别 3

年龄 age\_grade

职业 occupation\_grade

用户星级编号 start\_grade

搜索日期 1

搜索小时 1

商品id 商品一级类目 cate\_0

商品二级类目 cate\_1

商品三级类目 cate\_2

商品品牌 brand

商品属性 3 商品城市id city\_num

价格等级 price\_level 销量等级 sale\_level

展示次数等级 pv\_level

收藏次数等级 collected\_level

店铺id 1 店铺评价数量等级 review\_num\_level

好评数量等级 positive\_rate\_level

服务态度评分 float 0~1

物流服务评分 float 0~1

描述相符评分 float 0~1

当前是否工作日

小时划分 早晨/下午/晚上

展示页编号 int

查询词属性 one-hot cate\_num\*prop\_num

- 时间聚合特征(3级)

user\_id和context关联得到的特征

用户当日的搜索次数 1

用户当前小时的搜索次数 1

用户当日搜索商品次数 1

用户当前小时搜索商品次数 1

用户当日搜索店铺次数 1

用户当前小时搜索商店次数 1

用户当日搜索该类目次数 3

用户当前小时搜索该类目次数 3

用户当日搜索该关键词次数(hard)

用户当前小时搜索该关键词次数

各个搜索次数的统计值，最大、最小、均值（2级）

- 时间差特征(3级) 买完应该不会再搜索了

用户当次搜索距离当天第一次搜索该商品时间差 1

用户当次搜索距离当天最后一次搜索该商品时间差 1 用户当次搜索距离当天第一次搜索该商铺时间差 1

用户当次搜索距离当天最后一次搜索该商铺时间差 1 用户当次搜索距离当天第一次搜索该品牌时间差 1

用户当次搜索距离当天最后一次搜索该品牌时间差 1

用户当次搜索距离当天第一次搜索该类目时间差 1

用户当次搜索距离当天最后一次搜索该类目时间差 1

- 总值统计(2级)

用户买入总数 1

商品买入总数 1

商户买入总数 1

品牌买入总数 1

类目买入总数 1

- cvr平滑特征（搜索转化率 3级）

用户这天之前的搜索转化率平滑 1

店铺这天之前的搜索转化率平滑 1

商品这天之前的搜索转化率平滑 1  
商品品牌这天之前的搜索转化率平滑 1  
商品类目这天之前的搜索转化率平滑 3

- 相对特征 2级

商品相对同类目平均价格差 3

商品相对同品牌平均价格差 1

商品相对同类目平均销量差 3

商品相对同品牌商品平均销量差 3

- 模型

FFM+XGB+LGB