

天猫复购预测之挑战

特征字段对照表

用户特征处理（将数据集以用户维度做聚合）

- u1: 用户行为交互数量
- u2: item_id uv总数
- u3: cat_id uv总数
- u4: merchant_id uv总数
- u5: brand_id uv总数
- u6: 用户第一次和最后一次与店铺交互行为的时间间隔(按照小时)
- u7: 用户单击
- u8: 用户添加到购物车
- u9: 用户购买
- u10: 用户添加到收藏夹

商家特征处理（将数据集以商家维度做聚合）

- m1: 商家被交互的行为
- m2: user_id uv总数
- m3: item_id uv总数
- m4: cat_id uv总数
- m5: brand_id uv总数
- m6: 表示单击
- m7: 表示添加到购物车
- m8: 表示购买
- m9: 表示添加到收藏夹
- m10: 商家的老用户数

用户与商家关联特征处理（将数据集以用户、商家维度做聚合）

- um1: 用户对特定商家的交互数量
- um2: item_id uv总数
- um3: cat_id uv总数
- um4: brand_id uv总数
- um5: 表示单击
- um6: 表示添加到购物车
- um7: 表示购买
- um8: 表示添加到收藏夹
- um9: 用户第一次和最后一次与店铺交互行为的时间间隔(按照小时)

特征衍生

- r1: 用户维度聚合的购买点击比
- r2: 商家维度聚合的购买点击比
- r3: 用户与商家维度聚合的购买点击比

用户复购相关性分析

- 总体来看，商家相关的特征属性对用户复购率的影响最大，其次是用户特征。
- 商家特征中，商家的UV量越大，对用户的复购影响越大；商品的品类和品牌也对用户的复购起到一定的影响
- 用户特征中，用户对特定的商品和特征商铺比较敏感，这两项对用户的复购影响较大
- 从购买点击比对用户复购的影响看出，商家的特征属性要比用户的个人属性更重要
- 用户年龄与性别对于复购的影响较小

店铺通过各种营销手段吸引来更多的用户并形成购买行为，以及注重商品品类的区分，及品牌的营销，会更好地促进用户复购

lightgbm模型下的特征重要性

feature	importance
r2	2335
m2	2316
u4	2172
r1	2130
m3	1919
m4	1747
u6	1727
m8	1721
m7	1569
u3	1315
um9	1311
m9	1271
m2	1263
u9	1252
r3	1239
u2	1233
u4	1147
u7	1126
u10	1125
u1	1110
m1	1036
m10	1029

feature	importance
u5	1012
m6	830
um2	798
um1	743
um5	674
m5	478
um3	435
um7	368
um8	270
g_0	173
age_2	152
age_3	141
u8	137
um4	129
age_6	98
g_2	87
age_5	66
age_4	61
g_1	58
age_0	44
um6	35
age_7	18
age_8	11
age_1	0