

意愿度预估V1优化-添加渠道信息

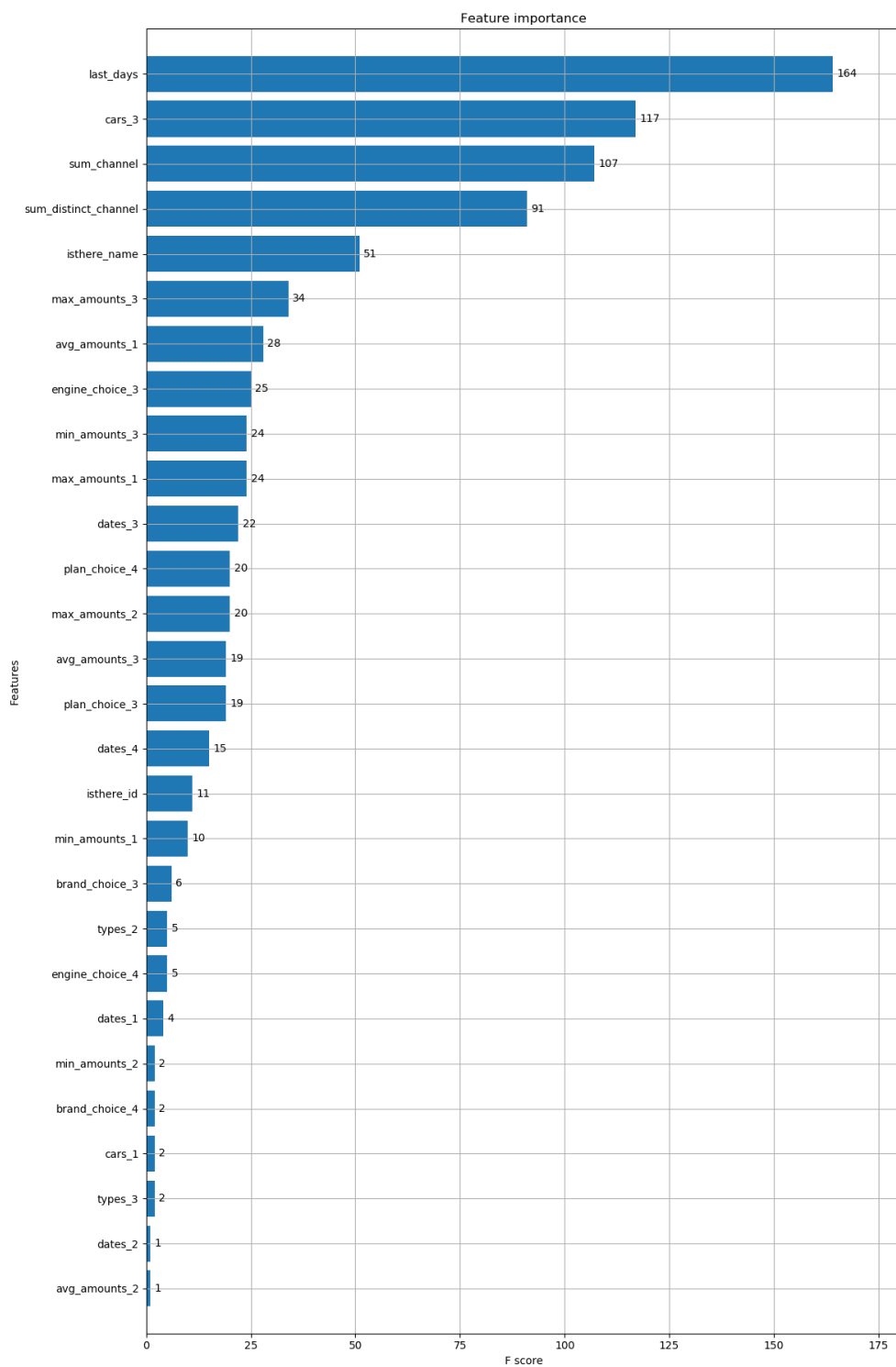
优化思路:

1. 添加channel_id字段(即来源渠道, 包括system, user_submit, reserve_submit, mis_import), 根据channel_id字段添加特征sum_channel(一个人出现在所有渠道中的次数), sum_distinct_channel(一个人出现在多少个不同的渠道)
2. 按总人群训练, 分渠道筛人群进行预测, 看不同渠道来源人群的平均表现

人群分析:

1. 渠道总人群数量: 560000左右 ,
654 cp_import
152 cscopy
97571 mis_import
17580 reserve_submit
432925 system
16662 user_submit
2. 渠道来源信息有浏览记录的数量: 163554 (单位: p_id)
3. 不同渠道人群分布(不同渠道人群可重叠, 单位: p_id):
 - a. system: 152006
 - b. user_submit: 11221
 - c. reserve_submit: 6707
 - d. mis_import: 7656
4. 不同渠道人群意愿情况分布(单位: 训练样本)
 - a. system: 预约 14795, 没预约 137211
 - b. user_submit: 预约 2337, 没预约 8884
 - c. reserve_submit: 预约 6168, 没预约 539
 - d. mis_import: 预约 795, 没预约 6861

添加特征后的特征重要性分析:



不同渠道人群预测的平均表现：

1. 对总人群进行正负样本1:6的采样后训练的结果：

AUC: 0.8809
 ACC: 0.9213
 Recall: 0.4940
 F1-score: 0.6342
 Precesion: 0.8857

2. 对system人群预测

- AUC: 0.7613
- ACC: 0.8905
- Recall: 0.6008
- Fl-score: 0.5164
- Precesion: 0.4528
- 3. 对user_submit人群预测
- AUC: 0.7220
- ACC: 0.7311
- Recall: 0.7065
- Fl-score: 0.5226
- Precesion: 0.4146
- 4. 对reserve_submit人群预测
- AUC: 0.7361
- ACC: 0.7279
- Recall: 0.7263
- Fl-score: 0.8308
- Precesion: 0.9703
- 5. 对mis_import人群预测
- AUC: 0.7434
- ACC: 0.8640
- Recall: 0.5912
- Fl-score: 0.4745
- Precesion: 0.3963
- 6. 对没有渠道来源信息的人群预测
- AUC: 0.6846
- ACC: 0.9857
- Recall: 0.3721
- Fl-score: 0.4894
- Precesion: 0.7147

对各渠道意愿分top100人群分析的结果:

共性:

1. 能查到搜索点击车型的浏览信息(区别于点击爆款,点击热门,进入详情页),且车型最大金额集中在30w-65w
2. 均有进入车型详情页的浏览行为
3. 来源渠道大于等于2的,在不同渠道中出现次数大于等于3的,有较大概率有高意愿

个性:

1. system:
上次登陆时间一定分布在30天以内,大多分布在15天以内
均填写过个人id与个人姓名
2. user_submit
上次登陆时间虽然有一些小的,但有大部分集中在30天上下
均填写过个人id与个人姓名
3. reserve_submit
上次登陆时间一定分布在11天以内,大多分布在6天以内
但是大部分人没有填写过个人id与个人姓名
4. mis_import
上次登陆时间大部分很小,有小部分集中在20天上下
但是大部分人没有填写过个人id与个人姓名

结论:

1. 根据宇鹏哥上次的触达实验,填写过个人id或者点击车型次数(cars_3)超过14的人群有较大可能有高意愿
2. 不同渠道在auc和recall上的平均表现无明显区别,在precision上区别较大,其中reserve_submit人群precision很高,这主要与该人群负样本太少有关
3. 根据这次实验,意愿高的人群有如下几个特征:填写过个人资料,登陆较为频繁,有自行搜索点击车型并进入详情页行为,来源渠道大于等于2且多次出现
4. 不同渠道人群的预测结果:(按意愿分从高到低排列)
 - a. system



data_all_pre_resu...t0710_system.xlsx

b. user_submit



data_all_pre_resu..._user_submit.xlsx

c. reserve_submit



data_all_pre_resu...serve_submit.xlsx

d. mis_import



data_all_pre_resu...0_mis_import.xlsx