

用户流失模型训练

赵海臣

训练 / 预测目标用户

- ▶ 目标用户：一年内有订单 / 浏览记录的用户
 - 若用户在一年内没有订单 / 浏览记录=>稳定死亡用户，无预测价值。

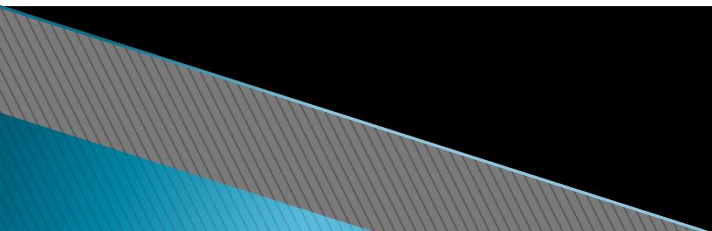
用户流失预测维度

▶ 用户维度

- 最近180天订单数量
- 最近180天订单总额
- 最近90天订单数量
- 最近90天订单总额
- 最近180-90天浏览次数
- 最近90-30天浏览次数
- 最近30天浏览次数
- 最近一次浏览时间
- 最近一次购买时间

用户流失评判指标

- ▶ 接下来90天内有浏览记录 => alive
- ▶ 接下来90天内无浏览记录 => dead



模型训练步骤

- ▶ 1. 设90天以前的某一天作为参考意义的“今天”；
- ▶ 2. 以“今天”作为标准计算当日的用户流失预测维度；
- ▶ 3. 以“今天”作为标准获得接下来90天的用户浏览记录，并对有浏览记录的用户标记“alive”；
- ▶ 4. 将用户分成(训练集/测试集) = (70%/30%)；
- ▶ 5. 将用户的预测维度输入决策树，并以“alive”作为用户的反馈数据，训练决策树；
- ▶ 6. 训练出决策树后，使用决策树对测试集进行预测，统计预测出“alive”准确性。
- ▶ 7. 参数调节：使用不同参数，寻找使“alive”预测准确性最高的参数。

最佳参数

▶ 决策树最佳参数：

◦ Best Param Found:

- maxDepth = 20
- maxBins = 64
- minInstancesPerNode = 1
- impurity = "gini"

测试结果数据

- ▶ 测试集用户预测准确性(alive/dead):
 - 以2017年1月9日为“今天”：82.4%
 - 以2016年10月10日为“今天”：84.9%
 - 以2016年7月15日为“今天”：81.3%
- 以2017年1月5日为“今天”：83.7%
 - 预测活跃用户/实际活跃用户：68.2%
 - 预测流失用户/实际流失用户：92.1%

用户动态行为与存活画像表

▶ 生成一个用户行为动态画像表

- 例子表: recommend.userAlivePortrait (现存, 可访问)
- 用户维度
 - 最近180天订单数量 – total_count
 - 最近180天订单总额 – total_amount
 - 最近90天订单数量 – recent_count
 - 最近90天订单总额 – recent_amount
 - 最近180-90天浏览次数 – recent_six_month_browse_count
 - 最近90-30天浏览次数 – recent_three_month_browse_count
 - 最近30天浏览次数 – recent_one_month_browse_count
 - 最近一次浏览时间 – recent_browse_date
 - 最近一次购买时间 – recent_order_date
- 用户流失预测: 1 / 0 – prediction
- 用户流失概率: [0%, 100%] – probability

The end