

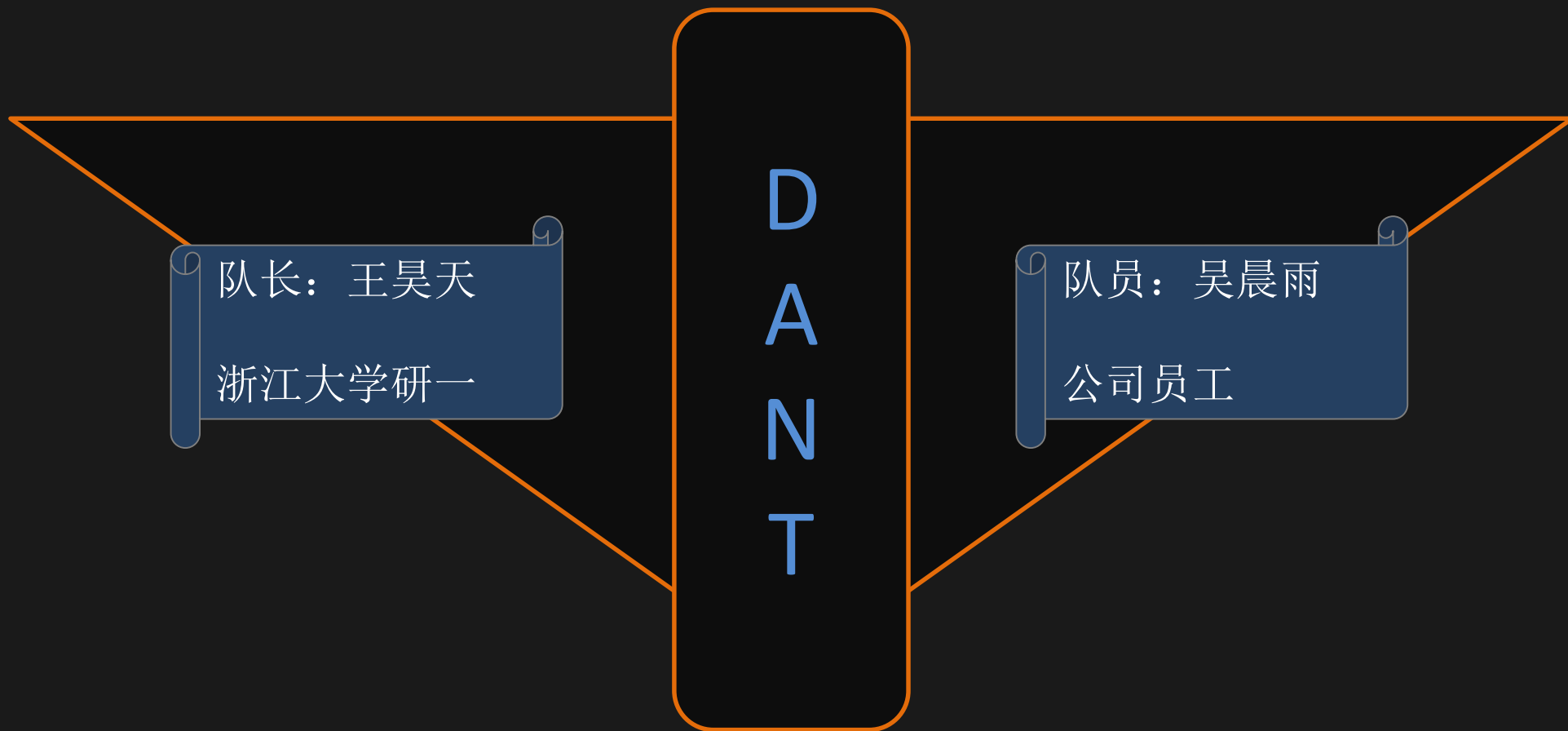
中国大数据算法大赛-用户购买时间预测

队伍名称：DANT

演讲者：王昊天

2018.07.19

团队介绍





CONTENTS

— 目录 —

01

赛题分析

02

数据划分

03

特征工程

04

总结回顾

01

赛题分析

赛题分析

赛题要求：根据用户一年内的订单、动作、评论信息判断用户下个月是否购买（分类问题）以及第一次购买的日期（回归问题）。

评价指标： s_1 决定 s_2 分数的上限，一般来说 s_1 提升 s_2 也会提升。

数据分析：根据用户短期内重复购买某商品，猜测商品有可能是牛奶或洗发水等生活必需品且损耗较快。



数据划分

数据划分

2017年 复赛数据

3月 4月 5月 6月 7月 8月 9月

2月 3月 4月 5月 6月 7月 8月

⋮

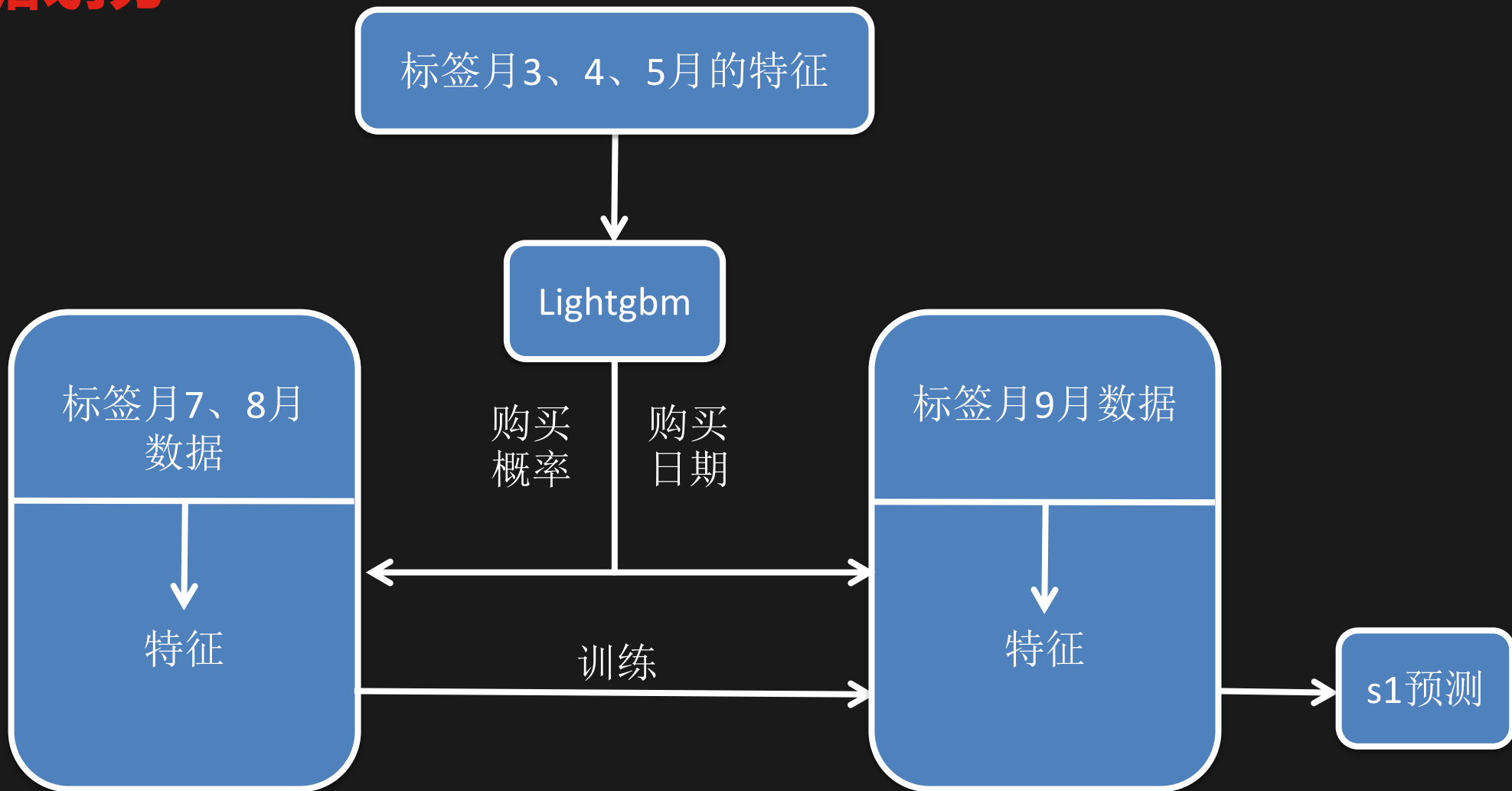
9月 10月 11月 12月 1月 2月 3月

2016年 初赛数据

6月 7月 8月 9月

5月 6月 7月 8月

数据划分





特征工程

特征工程



用户特征



统计特征



时间特征



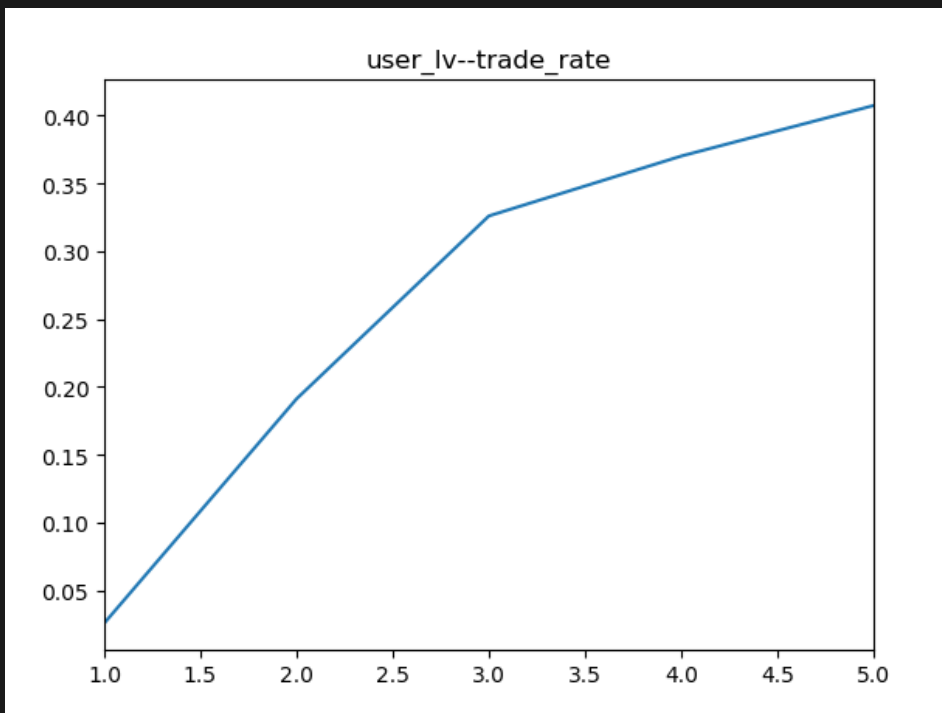
商品特征

用户特征

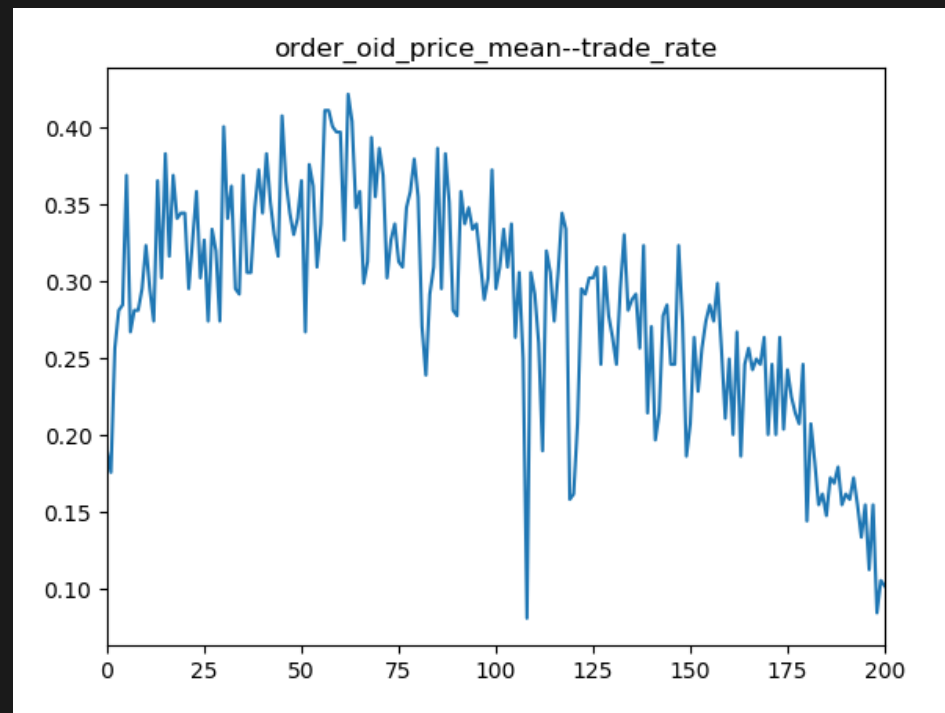
- 1、用户的性别、年龄、等级、地区。
- 2、用户每一单的价格、参数平均、最大最小值等。
- 3、用户购买、浏览目标商品的比例。

用户特征

用户等级与购买率



价格平均与购买率



统计特征

对全部种类商品和目标种类商品分别统计。

订单表

细粒度滑窗下的
不同商品数量
购买月份
购买次数
购买数量
购买日期
等

动作表

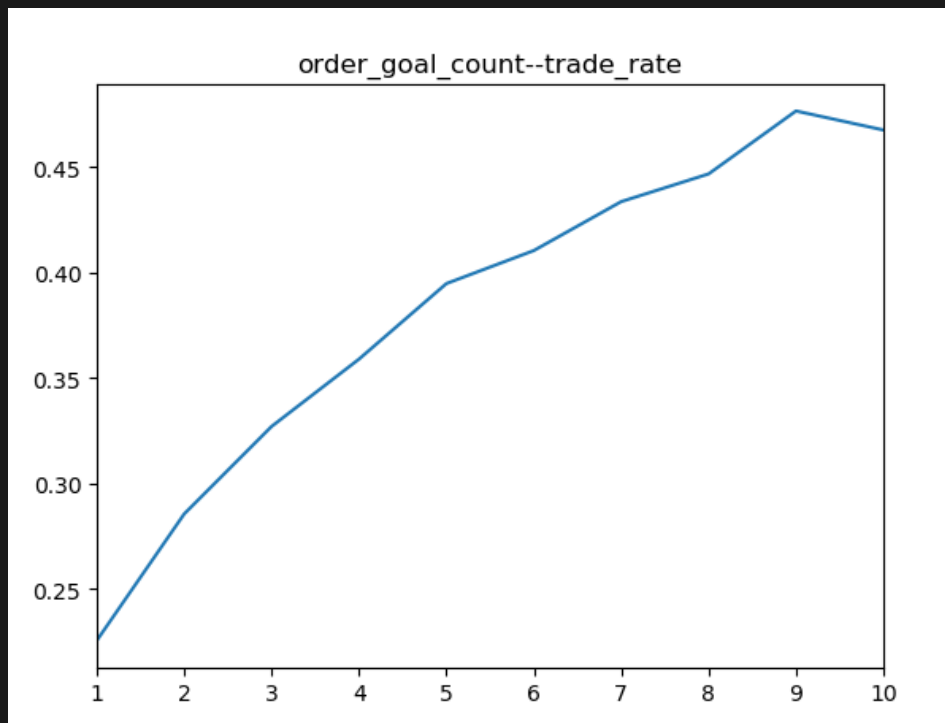
最后一次订单行为后5天的
浏览次数
浏览数量
浏览日期
收藏次数
对购买过的商品的浏览行为
等

评论表

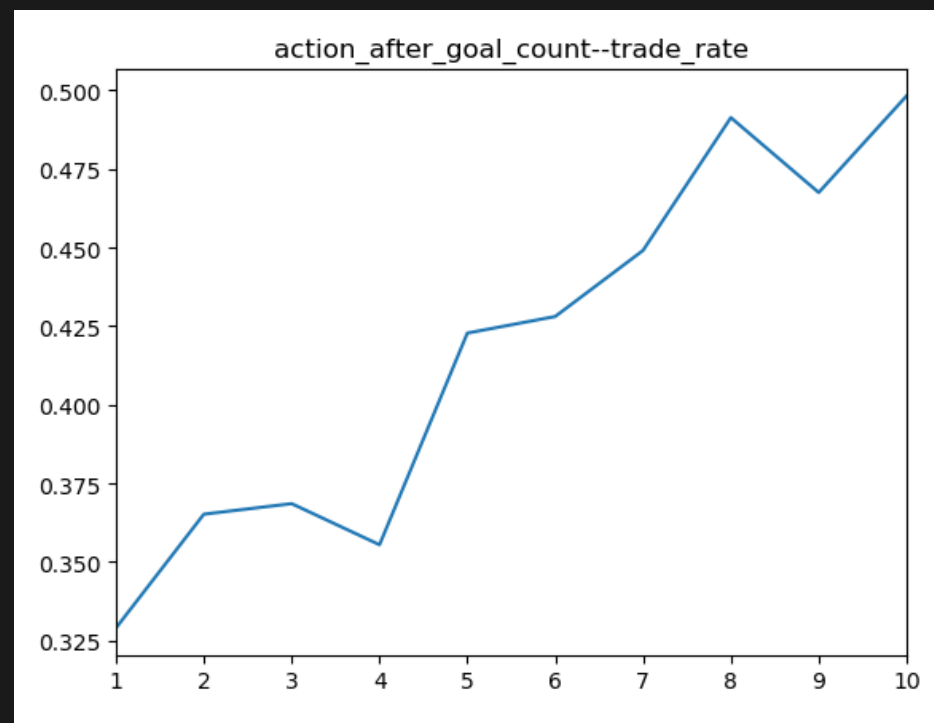
最后一次订单行为后的
评论次数
中评次数
差评次数。

统计特征

订单次数与购买率



浏览次数与购买率



时间特征

对全部种类商品、目标种类商品、30、101
种类商品分别统计。

订单表

上次购买日期

平均购买时间

最后一次购买时间

下次购买日期预测

购买时间差

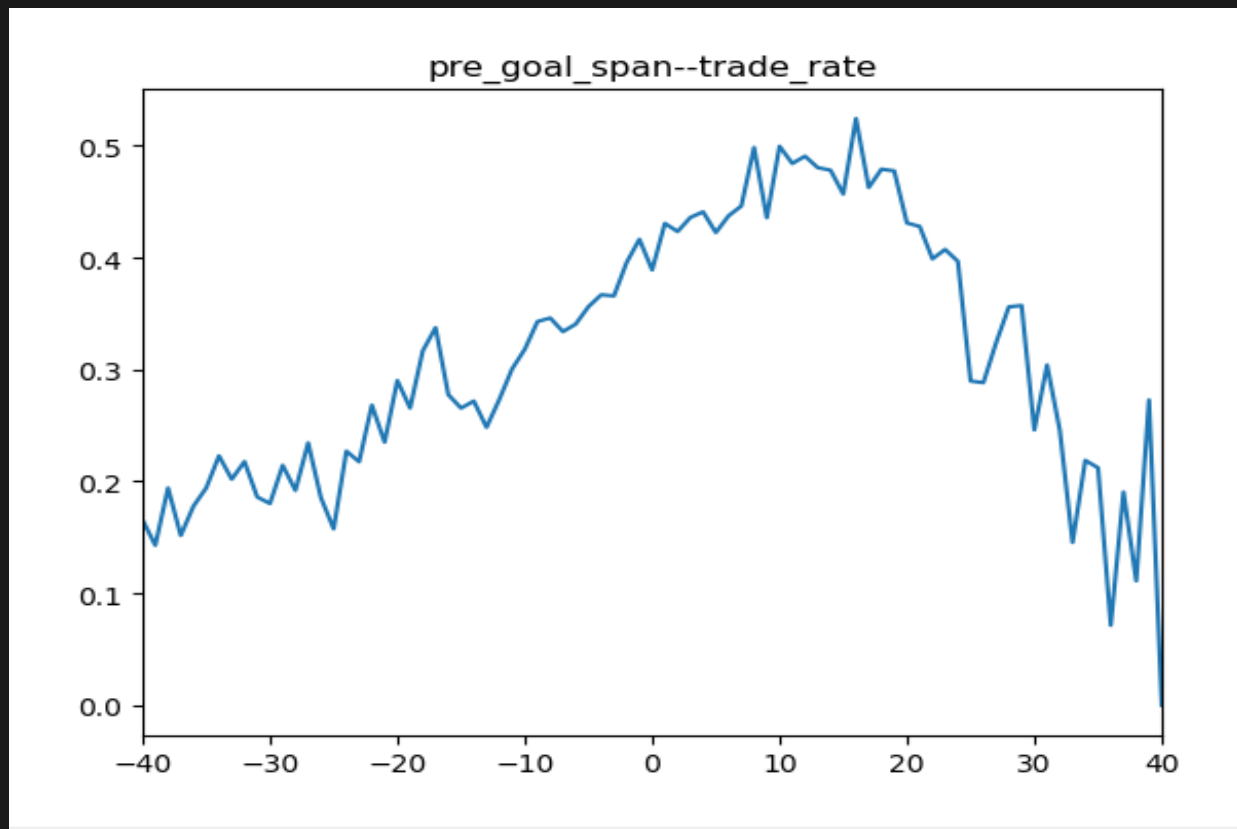
动作表和评论表

最后一次浏览日期

最后一次评论日期

时间特征

预测日期与购买率



商品特征

对全部种类商品、目标种类商品、30、101
种类商品分别统计。

订单表

上次购买日期

平均购买时间

最后一次购买时间

下次购买日期预测

用户购买该商品的次数

商品被用户购买的平均次数

商品特征

user_id	sku_id	o_sku_num	o_date	cate
22	29614	2	2017-05-17 00:00:00	101
22	29614	1	2017-06-18 00:00:00	101
22	81758	1	2017-07-12 00:00:00	101
22	29614	2	2017-07-23 00:00:00	101
			2017-08-19	

单个用户：平均购买时间、最后一次购买时间、连续购买的平均时间

所有用户：平均购买时间、最后一次购买时间、连续购买的平均时间

商品特征

用户	商品1	商品2	商品3	商品4
.....

对于S1，找到4个预测日期距离15号最近的商品信息。

对于S2，找到4个预测日期距离1号最近的商品信息。

04

回顾总结

回顾总结

不足：

- 1、前期构造了很多费时且没什么用的特征，由于时间原因也一直没有去整理，导致特征工程比较耗费时间。
- 2、商品与商品，用户与用户之间的关联性挖掘的不够充分。
- 3、对于s2，一个用户周几的下单偏好应该是有用的，但是没有构造出合适的特征。



THANKS