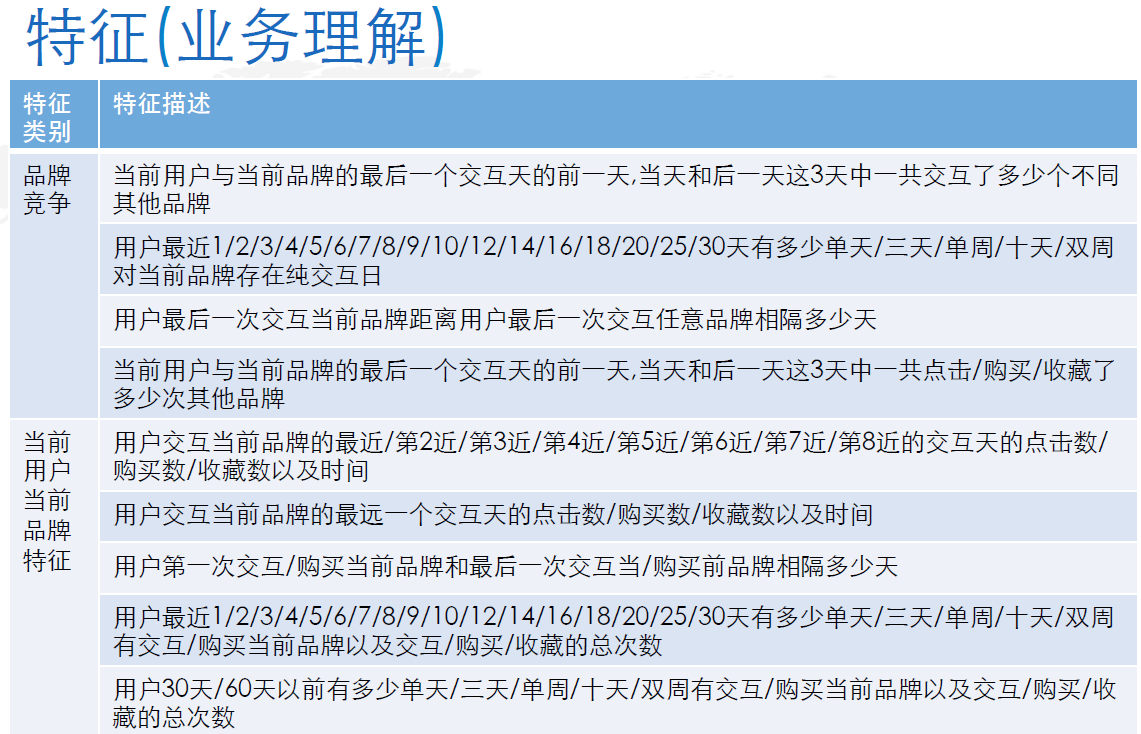
# 8th – 清水湾沙滩青年流浪者

## 特征抽取

在特征抽取方面，以业务理解为主，采取人工端只增不删的原则，对于稍有业务含义的特征均直接加入模型，对于关联性较大的特征之间并不做去重，以保证足够多的信息能被模型使用，加强模型的拟合能力。





# 7th - Bazinga

## 特征抽取

|  |  |  |  |
| --- | --- | --- | --- |
|  | Pair Features | User Features | Brand Features |
| Counting Features | action# | action# | action# |
|  | action day# | action day# | action day# |
|  | valid click# | distinct action brand# | distinct action user# |
|  | valid click day# | distinct first action brand# | distinct first action user# |
|  |  | action per brand per day# | action per user per day# |
| CVR Features | cvr (buy#, click#) | cvr (buy#, click#) | cvr (buy#, click#) |
|  | cvr (buy day, click day) | cvr (buy day#, click day#) | cvr (buy day#, click day#) |
|  |  | cvr (distinct buy brand#, distinct click brand#) | cvr (distinct buy user#, distinct click user#) |
| Ratio Features | ratio (click#, click day#) | ratio (click a#, click b#) X C52 | ratio (click a#, click b#) X C52 |
|  | ratio (buy#, buy day#) | ratio (buy a#, buy b#) X C52 | ratio (buy a#, buy b#) X C52 |
| Flag Features | action flag | action flag | action flag |
|  | other rules |  |  |
| Global Features | first / last action day | first / last action day |  |
|  | first / last buy day | first / last buy day | first / last buy day |
|  | active range length |  | frequent user ratio |

Counting特征指的是不同粒度上对行为数的聚合。

CVR特征指的是转化率特征。

Ratio特征是指同一种行为在不同粒度上的比例特征。

Flag特征是指一种0-1标志特征，表示是否存在这种状态，可以用来融合进各种规则。

Global特征是指与时间区间无关的全局特征。

# 5th - Tyche团队

## 特征抽取



# 4th – 数据心跳

# 3rd：Fly402

## 特征抽取

训练样本的基本形式是一对有交互的用户品牌对，我们对这样的用户品牌对提取特征。首先，从总的分类上分为用户的特征、品牌的特征、用户对这个品牌的特征；其次，考虑用户行为的周期性，从时间维度上分为最近一个星期、半个月、一个月等不同时间段的特征；然后，根据这次比赛给定的四种行为（点击、购买、收藏、购物车），从类型维度上去提取特征。在多维度考虑的基础上，主要提取的是点击数、天数、订单数这样的特征。

这次比赛的目的是由给定的数据去促进业务的提高，从数据到业务需要算法以及特征体系的驱动，在特征抽取的过程中，最本质的一点是要去让数据“贴合”业务，反过来就要求我们从业务指标出发提取特征。参照一些电商分析网站以及自己对业务的理解，诸如用户购买力、品牌购买成本、转化率、用户对品牌的忠诚度等等指标都可以从给定的数据集得以体现。典型的，比如如何衡量品牌的购买成本，可以使用购买该品牌的用户数、回头客的数量、平均每个回头客的回头次数等等来体现这个指标。

特征抽取的最后一部分主要是在之前特征体系的基础上做一些补充，主要有：

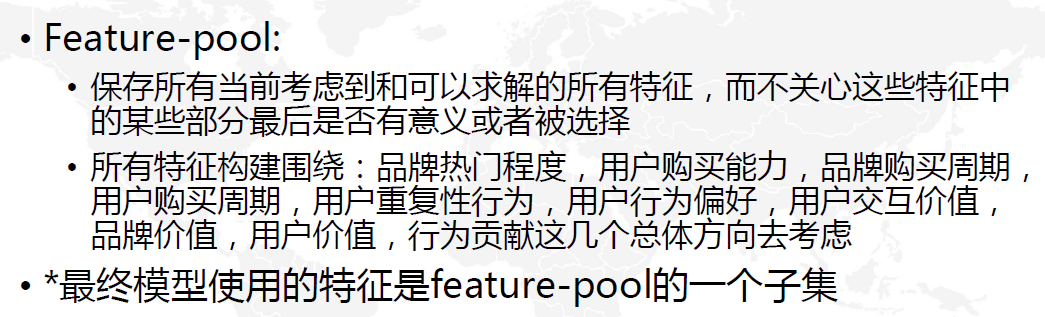
比值类特征。主要是一些基本特征的相除，这个一般是具有直观含义的，比如用户购买某个品牌的次数除以用户总的购买次数，这可以体现这个品牌在用户整个购买行为中的地位。

均值、方差类特征。主要是一些购买次数、活跃天数等统计值的平均值或方差，可以体现某些用户某些行为的平均水平或者波动情况。

策略。Season1的时候更多的考虑通过策略来预测，Season2的时候把一部分策略转化为特征，比如最后一次购买之后的点击次数等。

# 2nd：KLMJV

## 特征抽取



# 1st：Marvel

## 特征抽取

在特征提取过程中，我们是基于对业务的理解和对数据的分析来提取特征。特征分为四大类：行为特征、品牌特征、用户特征和交叉特征。

行为特征主要描述一个用户u对某个品牌b的操作历史，如操作次数、操作间隔等。行为特征提取过程中会考虑时间衰减。

品牌特征主要描述一个品牌的属性，如品牌的热度、销量、转化率。

用户特征则描述了用户的属性，如用户的活跃程度、用户转化率等。

交叉特征是基于前三类特征构造出来的新特征。例如用ub对的点击次数除以该用户u的点击次数，刻画了用户u对品牌b的偏爱程度。

