

# TYPE.1 count/ratio - count

按 {U, B, C, I, UB, UC, UI, BC} 分组

按monthly / whole period 统计

1. 购买（主体为用户）/被购买（主体为B、C、I）的次数
2. 购买（主体为用户）/被购买（主体为B、C、I）的金额
3. 购买/被购买发生的天数

共  $3 * 2 * 8 = 48$  个特征

# TYPE.1 count/ratio - product diversity

按 {U} 分组

按monthly / whole period 统计

1. 购买的unique的I的数量
2. 购买的unique的B的数量
3. 购买的unique的C的数量

按 {B, C} 分组

按monthly / whole period 统计

1. 包含的unique的I的数量

共  $3 * 2 * 1 + 1 * 2 * 2 = 10$  个特征

# TYPE.1 count/ratio - penetration

按 {B, C, I} 分组

按monthly / whole period 统计

1. 买过该品牌或品类或商品的不同的买家的人数

共  $1 * 2 * 3 = 6$  个特征

## TYPE.2 AGG feature - month AGG

针对TYPE.1 中所有的monthly特征，都可以进行此aggregation  
agg操作包含mean、std、max、median

共  $(48/2 + 10/2 + 6/2) * 4 = 32 * 4 = 128$  个特征

# TYPE.2 AGG feature - user AGG

按 {B, C, I} 分组

每组内先针对单个用户进行统计, 再进行aggregation

针对user统计

1. 总时间内发生购买的天数
2. 总时间内购买的次数
3. 总时间内购买的金额

agg操作包含mean、std、max、median

共  $3 * 3 * 4 = 36$  个特征

# TYPE.2 AGG feature – brand/category/item AGG

按 {U} 分组

每组内先针对单个B、C、I进行统计，再进行aggregation

针对B、C、I统计

1. 总时间内发生被购买的天数
2. 总时间内被购买的次数
3. 总时间被购买的金额

agg操作包含mean、std、max、median

共  $1 * 3 * 3 * 4 = 36$  个特征

## TYPE.3 last week / last month feature

将时间范围缩小，再进行TYPE.1 和TYPE.2 的特征统计，此时没有monthly特征和monthly agg特征

TYPE.1 特征数为原来一半  $(48 + 10 + 6) / 2 = 32$

TYPE.2 特征数为  $0 + 36 + 36 = 72$

共 $32+72=104$ 个特征

## TYPE.4 complex feature - trend

1. 基于TYPE.1 中的monthly feature, 一个monthly feature计算出一个trend
2. 基于TYPE.1 中的monthly feature, 算最后一个月与前几个月均值的偏差, 并用均值或者标准差normalize

共 $(48+10+6)/2 * 3 = 96$ 个特征

计算方法示例：假设monthly feature为前6个月的某一统计值(a1,a2,a3,...,a6), 那么有(1, a1), (2, a2, ), ... ,(6, a6)6个点, 根据这6个点可以用最小二乘法计算出拟合直线点斜率(slope), 该斜率即作为trend



# TYPE.4 complex feature – repeat feature

## 按 {B, C, I} 分组

统计repeat buyer的count/ratio

1. count : 至少两天购买同一I或B或C的人数
2. ratio : count / 购买该I或B或C的总人数

统计repeat day的count/ratio

1. count : 先每个user计算repeat buy day天数, 再求和
2. ratio : count / (先每个user计算buy day天数, 再求和)

## 按 {U} 分组

统计repeat {I, B, C}的count/ratio

1. count : 被该用户重复购买过的商品的数量 ==> 重复按天计算!
2. ratio : count / 该用户购买过的商品总数

统计repeat day of {I, B, C}的count/ratio

1. count : 先计算每个{I, B, C}的repeat buy day的天数, 再求和
2. ratio : count / (先计算每个{I, B, C}的buy day的天数, 再求和)

共  $3 * 4 + 1 * 4 * 3 = 24$  个特征

# TYPE.4 complex feature - market share

按{BC}分组

计算

1.  $N_{BC}/N_B$

2.  $U_{BC}/U_B$

3.  $N_{BC}/N_C$

4.  $U_{BC}/U_C$

共 4 个特征

# TYPE.4 complex feature – user-brand/category similarity

## 按{UB, UC}分组

以UB为例，某一brand中的不同C占比分别为(C1, C2, C3)

占比有两种算法：

1. 按购买数量算
2. 按购买的user数量算

某一user购买不同C的数量分别为( $N_{C1}$ ,  $N_{C2}$ ,  $N_{C3}$ )

similarity =  $\text{sum} (C1 * N_{C1}, C2 * N_{C2}, C3 * N_{C3})$

或 =  $\text{max} (C1 * N_{C1}, C2 * N_{C2}, C3 * N_{C3})$

或 =  $\text{avg} (C1 * N_{C1}, C2 * N_{C2}, C3 * N_{C3})$

共  $2*2*2(\text{max,avg}) = 8$  个特征