Feature engineering

1. **Interactive** 每个商家的最相似的5个商家(统计商家之间的共同用户) 5+1

取当前用户对此商家(相似商家)的行为特征：点击数 收藏数 添加购物车数 购买数 6\*4

当前用户对这6个商家 双十一的四种行为比例 4\*6

1. **用户**

重复购买的比例 1

在双十一之前重复购买比例

年龄 1

性别 1

行为天计数 1

平均每天行为因子（4种操作加权和） 1

双十一的四种行为比例 4

是否新用户

1. **商家**

所有用户在此商家(相似商家)重复购买的比例 1\*6

对此商家的四位行为数 点击数 收藏数 添加购物车数 购买数

双十一之前重复购买比例

双十一当天的比例

老用户比例

1. 统计类
   1. 计数类：、行为对象计数等；
   2. 比值类：平均每天行为、回购率、每月行为占比等；
   3. 生命周期类：用户首次/最后行为日、生存周期
2. 离散特征
3. 为什么要设计这套特征？为了使ID类完全个性化，用在LR上效果特别好（线性模型，L1正则，可实现效果好）。
4. 离散的方法，是将类别|ID可取的值展开成多维稀疏向量，只有对应的部分标1，其余为0.我们离散的特征为：
   * 用户的年龄和性别；
   * 当前用户在当前商家购买的商品ID、商品类别ID、品牌ID；
   * 当前用户在当前商家重复购买的品牌ID。

作者：littlekid  
链接：https://www.jianshu.com/p/c57ea5f99119  
來源：简书  
著作权归作者所有。商业转载请联系作者获得授权，非商业转载请注明出处。