

京东JData算法大赛: [JData](#)

代码说明

- 这是[JData用户下单预测比赛](#)的User Model(用户模型)解决方案

背景与方案说明

- 用户对商品的点击或者下单，体现了极强的用户兴趣倾向，对它的刻画可以用于最终推荐结果展示与搜索结果的排序。
- 在这种场景下的推荐与搜索个性化排序，解决方法是构建二分类模型，判断一个user是否会在一个item上发生点击/下单(训练集根据实际用户行为设为1或者0)，在新的场景下，使用二分类模型可以判断[user_geature, item_feature]作为输入时，输出的概率p，根据概率大小进行推荐和排序。
- 典型的建模方式是：
 - 第①种，以 [用户特征向量, 商品特征向量]的拼接为输入x，以用户是否在该商品上发生点击/下单(这里是下单)为标签y，构建二分类模型
 - 第②种，因为 用户*商品 的对数量非常大，我们也可以单独构建用户模型，商品模型，即在上一种建模方式中，只用用户特征或者商品特征，再考虑聚合。
- 这里的实现方式，是构建User Model，对用户进行建模

目录说明

- code: 数据处理与建模代码文件
- cache: 中间过程的缓存数据
- data: 原始数据文件
- model : 最终生成的user

代码使用说明

- utils.py: 工具函数
- gen_features.py: 数据预处理，提取特征，产生训练集和测试集
- user_modelling.py: 代码运行入口，训练模型，产生结果

特征介绍

- 暂略