## 数据假设

根据题目介绍，用户分析的数据来源与两个表格，即已知的商家名单文件Data\_Seller\_Product.csv和半年时间内每一所有买家的行为数据文件Data\_Action\_20160101.csv（假设具体日期为2016年1月1日至2016年6月30日），最终结果文件保存在Data\_Seller\_Product\_User.csv文件中，各文件内数据格式如下：

（1）、Data\_Seller\_Product.csv

|  |  |
| --- | --- |
| seller\_id | 商家ID |
| product\_id | 商品ID |

说明：一个商家ID（seller\_id）可以对应多种商品ID（product\_id）。

（2）、Data\_Action\_20160101.csv

|  |  |
| --- | --- |
| user\_id | 用户ID |
| product\_id | 商品ID |
| s\_num | 浏览数量 |
| c\_num | 收藏数量 |
| r\_num | 推荐数量 |
| b\_num | 购买数量 |
| Date | 日期 |

（3）、Data\_Seller\_Product\_User.csv

|  |  |
| --- | --- |
| seller\_id | 商家ID |
| product\_id | 商品ID |
| user\_id | 用户ID |

说明：推荐结果中product\_id-user\_id数据对是唯一的，会为每种商品（product\_id）推荐20个（user\_id）潜在的消费者。

## 解决方案

### 概述

#### 1.1、问题分析

为商家推荐一定数量的用户，便于其进行针对性的营销活动。可以根据过去半年内的历史行为数据，推测用户在接下来一段时间（取10天）内购买某个商品的可能性大小，选取可能性前N（20）的用户进行推荐，对每个商家的每种商品的推荐用户进行汇总，得到每个商家所需要的推荐用户名单。本方案的目标是根据用户对商品历史行为数据来预测用户在未来一段时间（10天）内购买该商品的概率。

问题转换成2分类问题，从用户行为数据表（Data\_Action\_20160101.csv）中构建特征，同时提取训练数据集和测试数据集，训练模型。

#### 1.2、整体思路

以一个月为周期，