### 赛题介绍

本赛题适用于season1 和season 2的比赛，season 3 为线上赛（season 2 赛后公布题目）。   
在天猫，每天都会有数千万的用户通过品牌发现自己喜欢的商品，品牌是联接消费者与商品最重要的纽带。本届赛题的任务就是根据用户4个月在天猫的行为日志，建立用户的品牌偏好，并预测他们在将来一个月内对品牌下商品的购买行为。

#### 我们会开放如下数据类型：

|  |  |  |
| --- | --- | --- |
| **字 段** | **字段说明** | **提取说明** |
| user\_id | 用户标记 | 抽样&字段加密 |
| Time | 行为时间 | 精度到天级别&隐藏年份 |
| action\_type | 用户对品牌的行为类型 | 包括点击、购买、加入购物车、收藏4种行为  (点击：0 购买：1 收藏：2 购物车：3） |
| brand\_id | 品牌数字ID | 抽样&字段加密 |

用户对任意商品的行为都会映射为一行数据。其中所有商品ID都已汇总为商品对应的品牌ID。用户和品牌都分别做了一定程度的数据抽样，且数字ID都做了加密。所有行为的时间都精确到天级别(隐藏年份)。

### 评估指标

我们希望参赛队的预测, 预测的品牌准确率越高越好，也希望覆盖的用户和品牌越多越好，所以用最常用的准确率与召回率作为排行榜的指标。

准确率：http://gtms02.alicdn.com/tps/i2/T1dF1ZFDNXXXX_722j-210-42.png   
注：   
N 为参赛队预测的用户数   
pBrandsi为对用户i 预测他(她)会购买的品牌列表个数   
hitBrandsi对用户i预测的品牌列表与用户i真实购买的品牌交集的个数

召回率：http://gtms03.alicdn.com/tps/i3/T1lfuWFyVbXXc.5Bvj-181-42.png   
注：   
M 为实际产生成交的用户数量   
bBrandsi为用户i 真实购买的品牌个数   
hitBrandsi预测的品牌列表与用户i真实购买的品牌交集的个数

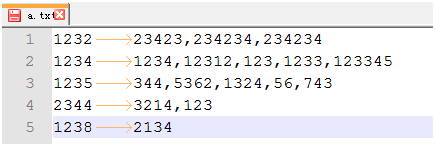
最后我们用F1-Score 来拟合准确率与召回率，并且大赛最终的比赛成绩排名以F1得分为准。   
http://gtms04.alicdn.com/tps/i4/T1Bg9YFrXbXXclgXI9-123-45.png

### 不同赛季使用数据说明

#### Season 1赛季数据

1)  数据获取格式：   
     参赛者通过大赛官网报名成功后，可以直接下载数据到本地，直接使用。（下载地址在参赛者的个人中心）   
2)  数据说明：   
     提供的原始文件有大约4M左右，涉及1千左右天猫用户，几千个天猫品牌，总共10万多条的行为记录。用户4种行为类型(Type)对应代码分别为：  
     点击：0；购买：1；收藏：2；购物车：3   
3)  提交评分方式：   
     参赛者将预测的用户存入文本文件中，格式如下：   
     user\_id \t brand\_id , brand\_id , brand\_id \n   
     上传的结果文件名字不限(20字以内)，文件必须为txt格式。

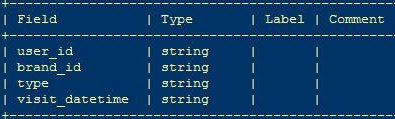
     如图：



     将预测结果文件上传至大赛官网结果提交入口即可。（提交结果入口在参赛者的个人中心）

#### Season 2赛季数据

1)  数据获取方式：   
     参赛者只能登陆阿里巴巴天池集群获取数据，且数据不能下载，所有数据分析，计算，提交评分都将在天池集群上完成。登陆天池集群方法，请留  
     意后续的指导文档和邮件通知。   
2)  数据说明：   
     提供的数据量，涉及千万级天猫用户，万级天猫品牌，时间跨度4个月的行为记录。   
     提供的训练数据在天池集群的表t\_alibaba\_bigdata\_user\_brand\_total\_1中，字段分别为：user\_id，brand\_id, type, visit\_datetime。如图所示



     用户4种行为类型(Type)对应代码分别为：   
     点击：0；购买：1；收藏：2；购物车：3   
3)  提交评分方式：   
     参赛者完成用户品牌偏好预测后，需要将结果放入指定格式的数据表（非分区表）中，要求结果表名为：t\_tmall\_add\_user\_brand\_predict\_dh，  
     包含user\_id和brand两列。其中brand字段为预测的偏好品牌ID，多个品牌之间用逗号(",")分隔。   
     例如：

