# 实时推荐关联分析算法功能设计

## 算法过程

### 关联分析 apriori算法

原理是浏览过或者购买过的该产品的人还看过什么。

将一个人的历史上一段时间浏览或者购买过的产品作为一个事务，利用apriori算法算支持度和置信度，排名，推荐前n个。

支持度：P(AUB)

置信度：P(A|B)

备选方案：同类产品热门排行

## 输入信息

**T\_DBP\_PASS\_USER\_VIEW(用户浏览行为记录表)**

说明: userAgent字段的值取决于用户是否有 mc\_code，如果用户有 mc\_code，useragent字段值就是mc\_code，如果没有，那就是session\_id

|  |  |  |  |
| --- | --- | --- | --- |
| family | column | value | 示例 |
| info | rowkey | 日期\_ID | 2000-03-23\_6D0F0B6F247E09A2B24B6C0C9A06335C |
| info | ip | 用户IP | 172.24.88.88 |
| info | userAgent | 用户标示, 会员登陆后记录会员mc\_code,临时用户则是session\_id | 4087449 |
| info | viewUrl | 访问地址 | http://club.jinjiang.com/hotel/ |
| info | eventType | 值是枚举类型（SEARCH,VIEW\_PROD,CLICK,BOOKING, BOOKING\_SUCCESS） | VIEW\_PROD |
| info | dataJson | 值是json格式字符串，如果是浏览行为，则值为产品ID | {"line\_id":"826431"} |
| info | viewType | 值是枚举类型：旅游产品（TRAVEL）、酒店产品（HOTEL） | TRAVEL |
| info | utrace | 字段记录用户cookie信息。 | 95C6280F56C9A84C1792FA0067FE936F |

**JJ000\_SIEBEL\_S\_LOY\_TXN\_TRIM（交易记录表）**

说明：X\_ORDER\_NUM (订单号)不唯一，存在多个ROW\_ID对应同一个X\_ORDER\_NUM的情况，也就是同一个订单（X\_ORDER\_NUM）可能会有多个人参加。每个人的价格不同，因而拆分出不同的消费记录。

|  |  |  |
| --- | --- | --- |
| family | column | value |
| info | ROW\_ID | 消费ID |
| info | TXN\_NUM | 消费编号 |
| info | X\_CARD\_NUM | 消费卡号 |
| info | X\_CARD\_TYPE | 消费卡类型 |
| info | STATUS\_CD | 消费状态 |
| info | TYPE\_CD | 消费类型 |
| info | SUB\_TYPE\_CD | 消费子类型 |
| info | PARTNER\_ID | 消费板块 |
| info | PROD\_ID | 消费产品 |
| info | TXN\_CHANNEL\_CD | 消费渠道 |
| info | PYMT\_TYPE\_CD | 支付方式 |
| info | BOOKING\_DT | 预定日期 |
| info | START\_DT | 开始日期 |
| info | END\_DT | 结束日期 |
| info | TXN\_DT | 消费日期 |
| info | PROCESS\_DT | 处理日期 |
| info | X\_ROOM\_NIGHTS | 间夜数 |
| info | AMT\_VAL | 可积分金额 |
| info | ORIG\_AMOUNT | 里程换算金额 |
| info | X\_RMB\_AMOUNT | 人民币金额 |
| info | X\_TOTAL\_AMOUNT | 总消费金额 |
| info | X\_ORDER\_NUM | 订单编号 |
| info | X\_PMS\_NUMBER | PMS编号 |
| info | X\_GROUP\_CODE | 团代码 |
| info | X\_ACCRUAL\_COMP | 合作伙伴卡类型 |
| info | X\_PARTNER\_CARD | 合作伙伴卡 |
| info | X\_SOURCE\_PARTY | 兑换积分来源板块 |
| info | X\_PRODUCTID/X\_REDEEM\_PROD | 礼品ID |
| info | X\_HOTEL\_ID | 酒店ID |
| info | X\_HOTEL\_NAME | 酒店名称 |
| info | MEMBER\_ID | 会员ID |
| info | X\_CONTACT\_ID | 联系人编号 |
| info | COMMENTS | 备注 |
| info | CANCELLED\_TXN\_ID | 取消交易ID |
| info | ACCRUALED\_VALUE | 定级积分 |

## 输出信息

大数据平台计算：输出产品关联结果表

旅游线路关联性分析结果表：

TB\_REC\_CH\_TRAVEL\_LINE\_RELATIVITY

rowkey: line\_id

|  |  |  |  |
| --- | --- | --- | --- |
| family | columnname | 示例 | columnvalue |
| info | line\_id | 10001 | 相应线路ID |
| info | Time | 20141224112900 | 时间 |
| line | 具体的line\_id | 0.7 | 匹配度 |