旅游实时推荐方案

# 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

接口开发：rowkey lineid

如果不够，需要从热门排行取，补全要求的数量

输入：lineid 产品数量

返回lineid的list

# 实时推荐

## 算法

1. 离线算好产品之间的相似度，得到最相似的m个产品：离线计算，结果存放到hbase里，以备后用。
2. 线上根据用户的实时浏览行为计算获取用户的实时兴趣，用用户的兴趣向量与用户浏览过的产品的相似产品进行匹配。

## 旅游产品的特征向量表示

一个旅游产品用一个n为0-1向量a表示，在旅游产品标签的基础上，增加目的地、逗留时间、价格区间等。

旅游产品标签：根据文本挖掘的结果，取前m（暂定为5）个，查看标签有无，取值为1。

目的地：大洲-国家-省份-景点 所有维度拉平

线路A：去阿尔巴尼亚的线路

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 中国 | 澳洲 | 亚洲 | 欧洲 | 非洲 | 美洲 | 山东 | 上海 | 河南 | 。。。。 | 阿尔巴尼亚 | 8000~100000 |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |

逗留时间分段：1~2 3~5 5~7 8~

价格区间：c(0,500,1500,8000,15000,30000,Inf)

：

金棕榈的线路信息：酒店星级等 要看哦 机票（直飞还是转机）

自由行 跟团游

还有什么维度？开动脑筋。。。。。。。。。。。。。。。。。。。。。

线路价格

select l.id,min(p.sale\_price) sale\_price

from t\_tbp\_group\_price p,t\_tbp\_group g,t\_tbp\_line l

where g.id = p.group\_id

and g.line\_id = l.id

and p.category = '基本价'

and p.type='团费'

and g.status = 'ON'

and g.current\_status = 'ACTIVATE'

and g.plannum> (g.paidinnum+g.reservenum)

and g.depart\_date> (g.advanceclosedso + sysdate)

and l.status ='ON'

group by l.id;

## 物品相似性计算----离线计算

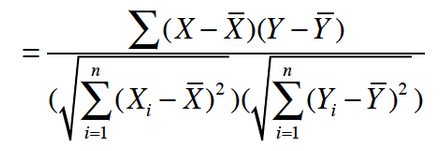
### 相似旅游产品计算

根据产品相似度，取*l*个相似的可预订线路，存放在hbase。

### 相似度计算方法

#### Pearson相似性

http://webdataanalysis.net/wp-content/uploads/2011/10/Pearson-Correlation-Coefficient.png



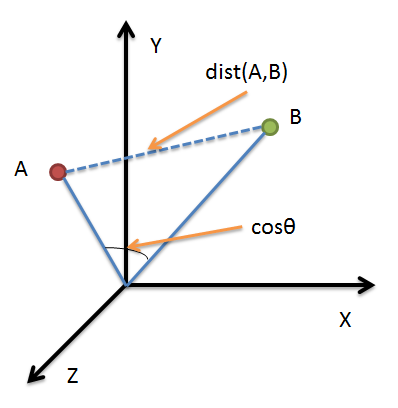
X (0,1,0,1,0)

Y (1,0,1,1,1)

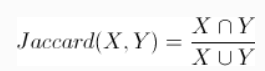


#### Cos相似性

http://webdataanalysis.net/wp-content/uploads/2011/10/Cosine-Similarity.png



#### Jacard相似度



### 本步骤的结果

产品A：产品b 产品c 产品d 产品e

## 实时计算推荐产品

### 用户向量表示

用户兴趣用一个n维向量b表示，维度与产品一样，不过不是0，1向量。随着时间衰减，衰减系数暂定为0.98.用户每浏览或者购买一个产品，将产品对应维度叠加到用户向量上。根据用户兴趣向量计算与浏览产品相似的产品的匹配度。



B为用户兴趣向量，为当前浏览产品与相似产品的相似度，为相似产品的特征向量。

# 实时推荐接口

输入:用户id（memberid）指定数量

输出：线路list