**Thinking1: 关联规则中的支持度、置信度和提升度代表什么,如何计算?**

答: 1.支持度:是个百分比,指的是某个产品组合出现的次数与总次数之间的比例,

支持度越高,代表这个组合出现的频率越大,计算方式如下

2.置信度:是个条件概率,指的是你购买了商品A,有多大概率会购买商品B,计算方式如 下:

3.提升度:商品A的出现,对商品B出现概率提升的程度,计算方式如下:

**Thinking2: 关联规则与协同过滤的区别?**

答: 关联规则可以从大量的过往交易数据中获取规则.她可以是会同时被购买的商品之间的

关联规则,也可以是按时间依次被购买商品的序列模型,因此关联规则是直接vong数据中挖掘潜在的关联,与个人的偏好无关,忽略了个性化的场景,

协同过滤分为基于用户的协同过滤,和基于商品的协同过滤,它主要是先找到相似的

物品或者相似的用户,然后再去进行个性化的推荐.

因此关联规则和协同过滤两者之间最大的区别就是后者属于个性化推荐,而前者忽略

了这一点.

**Thinking3: 为什么我们需要多种推荐算法?**

答: 因为推荐的方式和前提条件是不同的,比如冷启动用户就要使用协同过滤的方式进行推荐,而个性化推荐中又经常使用基于标签的推荐算法,再有基于内容的推荐和基于规则的推荐,甚至于基于人口统计信息的推荐.

**Thinking4: 关联规则中最小支持度、最小置信度该如何确定?**

答: 最小支持度、最小置信度都是实验出来的,要根据训练数据和场景的接受程度来定.这两个阈值给的越大,得到的有相关性的item就越好,但item的数量也越少.所以需要根据给出的推荐item的数量及item间的相关性来决定两个阈值的大小.

**Thinking5: 如何通过可视化的方式探索特征之间的相关性?**

答: 可以使用计算特征之间的Pearson系数的方式,展现在热力图上的方式探索特征之间的相关性.也可以使用Sperman秩相关系数的方式.

**Action1:针对MarketBasket数据集进行购物篮分析(频繁项集及关联规则挖掘)?**

答:

 

**Action2: 针对MarketBasket进行词云分析,可视化探索(Top10的商品有那些)?**

答:

