用户画像潜在需求指标设计方案

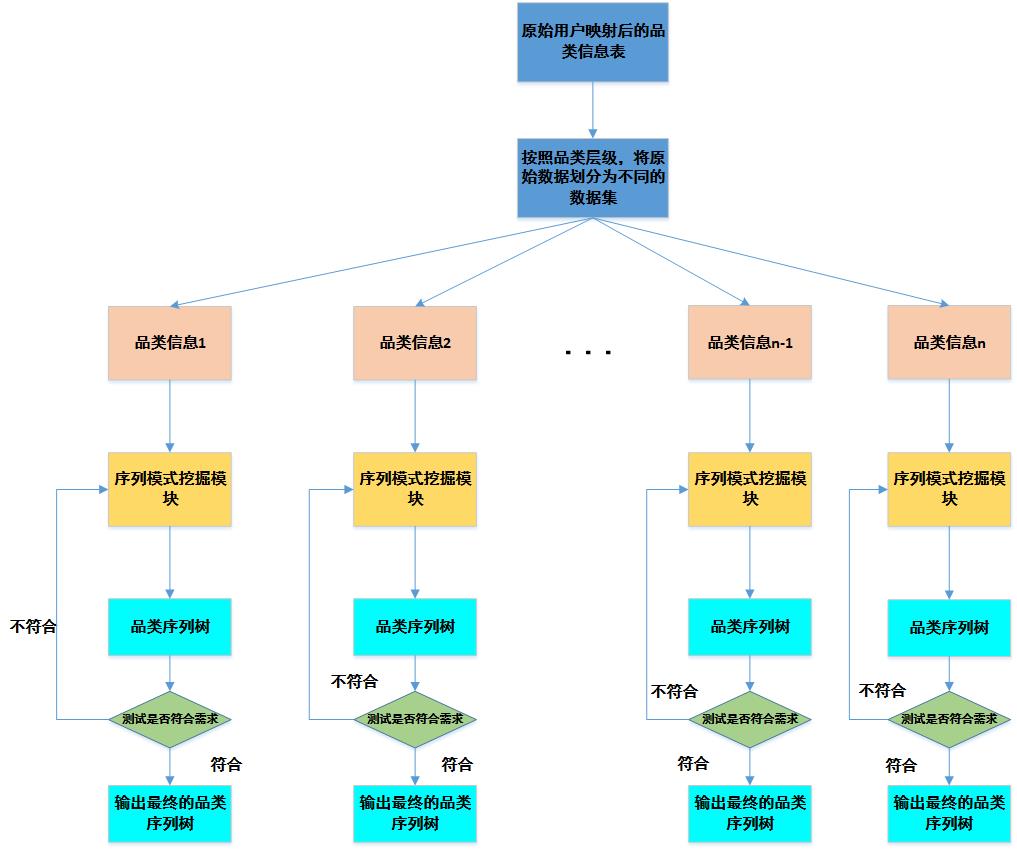
# 目标

根据用户的历史行为，构建品类潜在关系树。按照品类层次 ，给出各级品类预测树。

# 方案

潜在需求可以理解为购买某种商品之后，可能购买哪些商品。在用户画像潜在需求中体现了一种时间的概念，即购买了某种品类商品之后还可能哪些品类商品具有兴趣。因此根据上面的需求，我们既定的方案是按层级分类，使用序列模式挖掘算法获取品类的购买序列，最终按照层级关系形成相关的购买品类序列树。总体流程入分为训练流程，预测流程，更新流程。

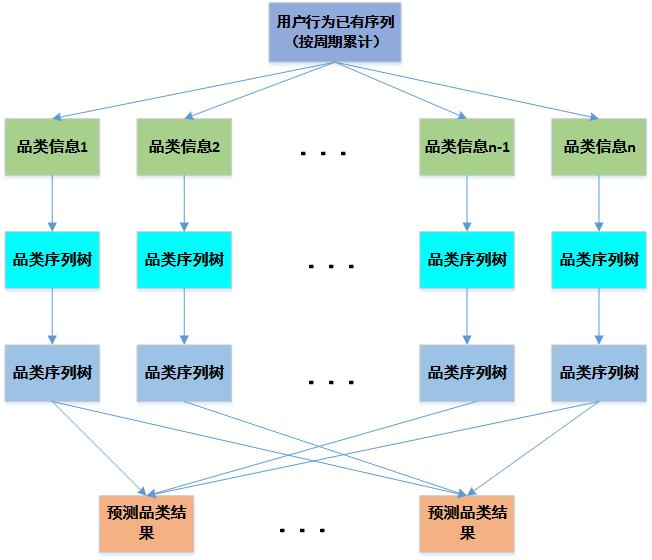
* 训练流程



在整个训练过程中，总体上分为以下几个步骤：

1. 数据集的划分。根据品类的1-5级进行划分，有多少个级别划分为多少个数据集。例如假如我们有一级品类20个， 二级品类40个， 三级品类50个，四级品类60个，五级品类100个，那么最终理论上划分的数据集的个数为270(20 + 40 + 50 + 60 + 100)
2. 品类序列树生成。 针对不同的数据集采用不同的支持度进行裁剪操作，可以设计不同的支持度，设计相关的裁决机制。对于生成的序列模式集，只选取相关的二项频繁序列模式集生成相关的品类序列树，使用对应的支持度作为相关的支持度，最终可以形成一个多叉树桩决策树。获取未来几天的用户行为数据，使用生成的多叉树桩决策树树，对于用户偏好的预测准确度进行测试，如果符合人为设定的阈值，输出相关的多叉决策树桩；否则，返回序列模式挖掘模块，重新设计决策机制以及选取频繁序列模式集的方法，生成新的多查封决策树桩。
3. 按照用户已有行为划分不同的原始训练集。例如按照浏览、下单、添加购物车或者收藏等划分相关的原始训练集，同时给出不同行为产生的对应权重，例如下单>添加购物车>收藏>下单等。

* 预测流程



在整个预测流程中，总体上分为以下几个步骤：

1. 用户行为已有序列：待测用户行为序列累计周期（按天、周或者月累计用户的行为序列），同时按不同行为累计用户的行为序列。
2. 划分品类信息：将待测用户行为序列按行为划分，分解不同的品类信息。
3. 根据不同的品类信息，划分到不同的品类树上进行相关的预测，给出对应的预测结果
4. 预测结果融合：将预测得到的相同结果融合起来，即将对应的相同结果的权重累加在一起，构成最终结果的权重值，最终产生对应的预测结果。