# Pratice

## 1-Thinking1：既然内容相似度计算简单，能频繁更新，为什么还需要协同过滤算法呢？

Answer：内容相似度计算简单，能频繁更新，且能解决物品冷启动问题。内容过滤算法ContentItemKNN在某些数据集上的准确率、召回率上性能较差，覆盖率、流行度上性能稍好。

很多时候内容过滤算法的精度比协同过滤算法差，但如果用户的行为强烈受某一内容属性的影响，那么内容过滤算法还是可以再精度上超过协同过滤算法的。如果能将这两种算法融合，一定能获得比单独使用这两种算法更好的效果。向量空间模型在内容数据丰富时可获得较好的效果。

就好比这样的场景：一个在线网站，用户的数量往往大大超过物品的数量，同时物品的数据相对稳定，因此计算物品的相似度不但计算量较小，同时也不必频繁更新。但我们往往忽略了这种情况只适应于提供商品的电子商务网站，对于新闻，博客或者微内容的推荐系统，情况往往是相反的，物品的数量是海量的，同时也是更新频繁的，所以单从复杂度的角度，这两个算法在不同的系统中各有优势，推荐引擎的设计者需要根据自己应用的特点选择更加合适的算法。

在非社交网络的网站中，内容内在的联系是很重要的推荐原则，它比基于相似用户的推荐原则更加有效。比如在购书网站上，当你看一本书的时候，推荐引擎会给你推荐相关的书籍，这个推荐的重要性远远超过了网站首页对该用户的综合推荐。可以看到，在这种情况下，Item CF 的推荐成为了引导用户浏览的重要手段。同时 Item CF 便于为推荐做出解释，在一个非社交网络的网站中，给某个用户推荐一本书，同时给出的解释是某某和你有相似兴趣的人也看了这本书，这很难让用户信服，因为用户可能根本不认识那个人；但如果解释说是因为这本书和你以前看的某本书相似，用户可能就觉得合理而采纳了此推荐。

　　相反的，在现今很流行的社交网络站点中，User CF 是一个更不错的选择，User CF 加上社会网络信息，可以增加用户对推荐解释的信服程度。（微信读书等）

## 2-Thinking2、你需要推荐系统么？哪些情况下不需要推荐系统？

Answer：需要，推荐可以带来更多的选择。货比三家，自己可以选择比之前更好的，某种程度上也是很节省自己的时间。

推荐系统的前提是需要足够的数据。到底要不要上推荐系统，如果仅仅从利益上来看，是一个关乎投入产出比的问题，搭建一个推荐系统的前期投入不小，你需要：组建团队、购置计算资源、积累数据、花费时间优化。这些都是很大的支出。--平台数据量小，增量数据小的公司可以不用考虑，成本大。

## 3-Thinking3、如果给一个视频打标签，视频中有音乐作为背景音乐，采用了NLP方式对内容自动打标签，可能存在什么问题？

Answer：导致推荐的内容出现偏差，原因是NLP的方式将背景音乐也处理成为特征向量。