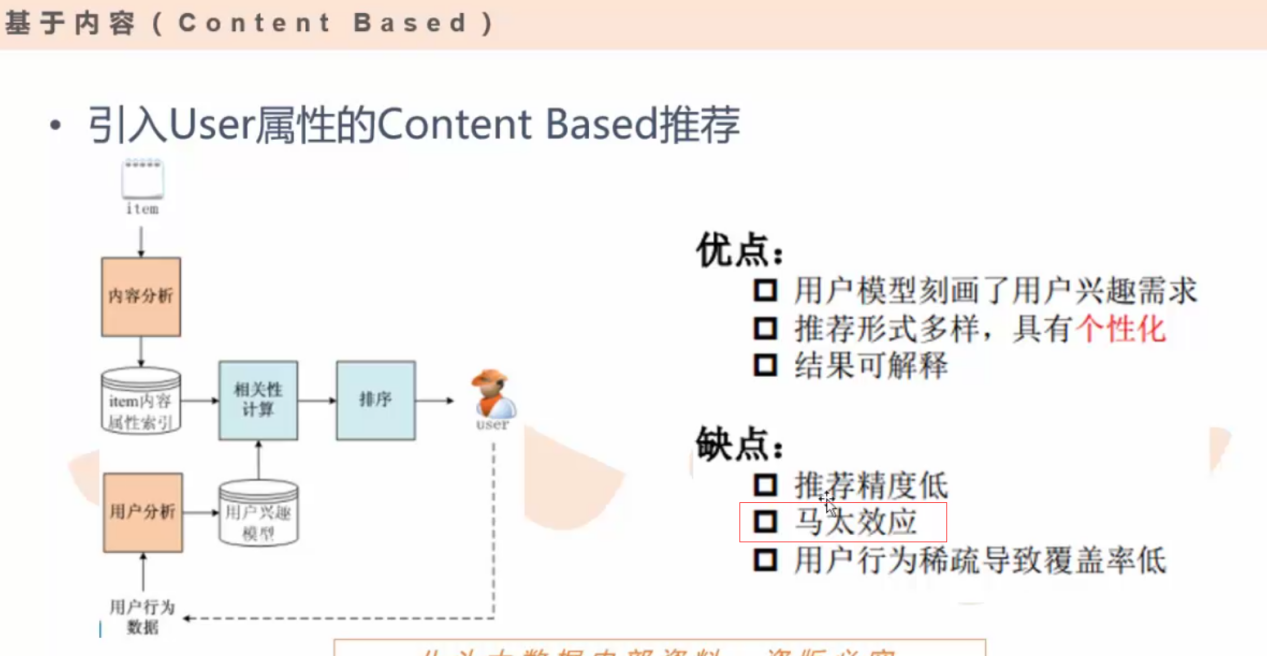
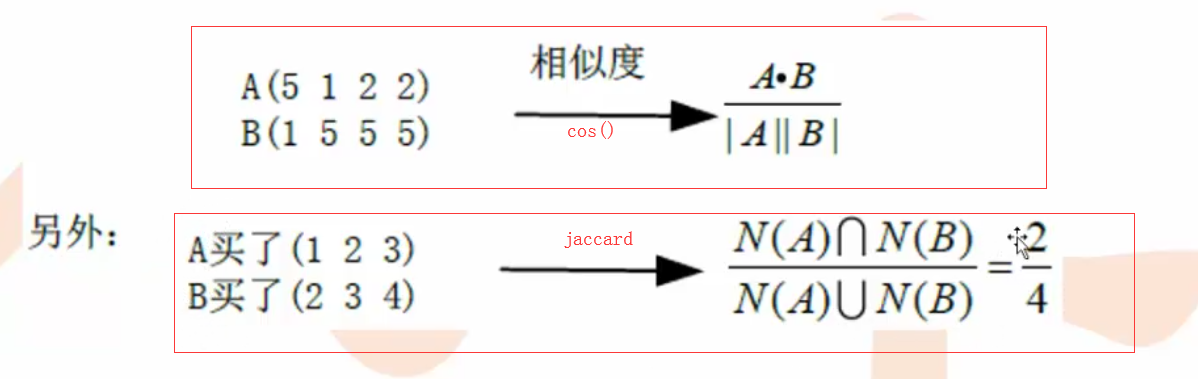
**目的给user推荐item**

1. 基于内容推荐content based

根据文本相关性。考虑词的共现状况。



相似度除了cos() 还有jaccard



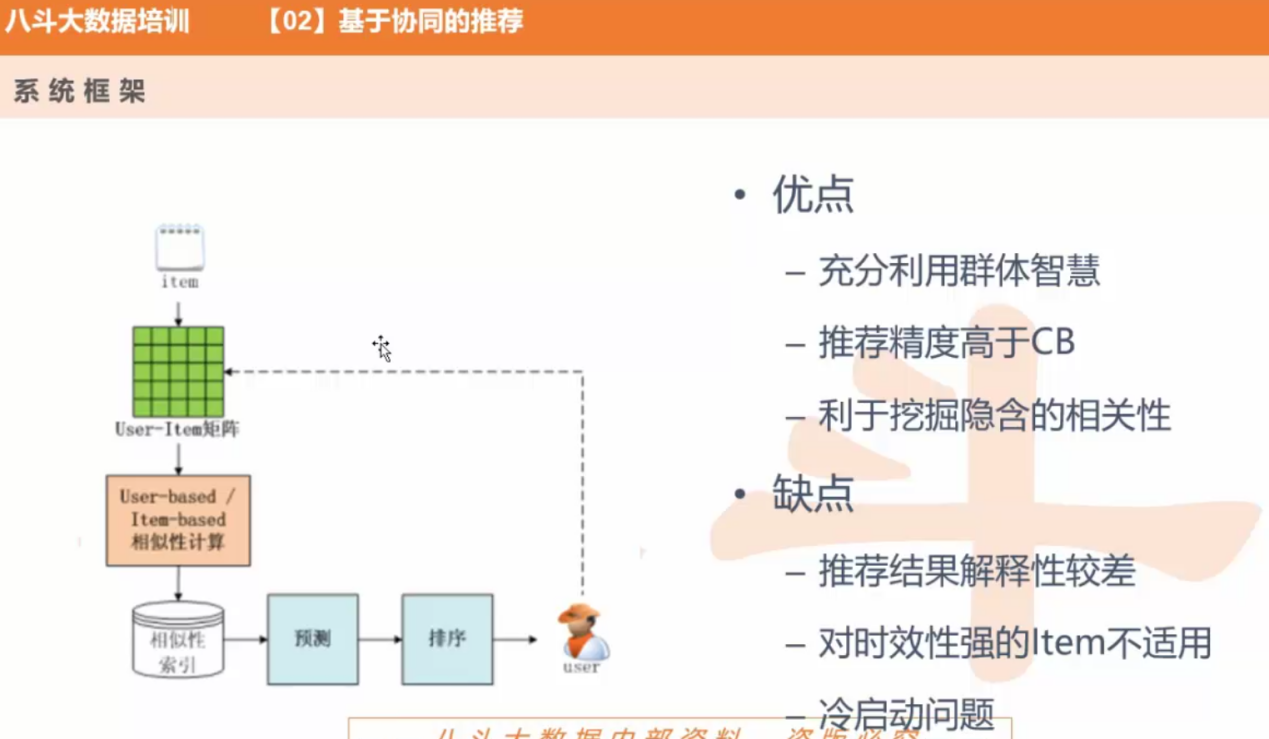
基于内容item的token分析

1. 基于协同过滤的推荐(collaborative filtering)

**基于用户行为**，本质上是graph model。考虑的是用户的共同行为。

内容推荐算是协同推荐的子集

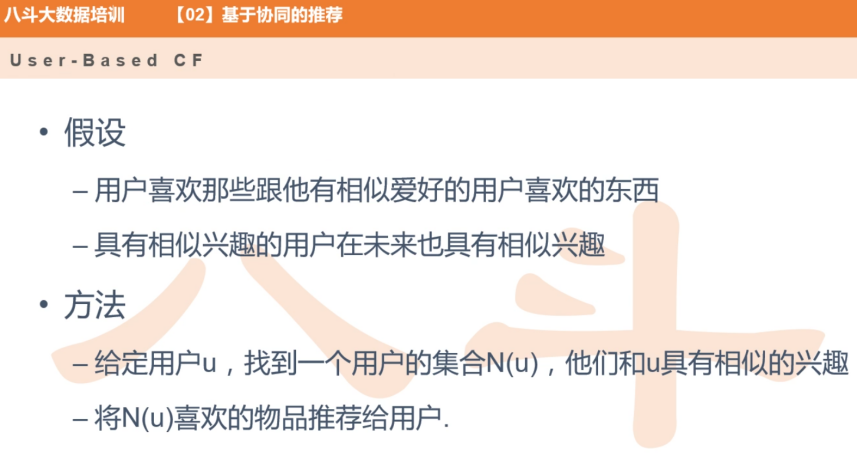


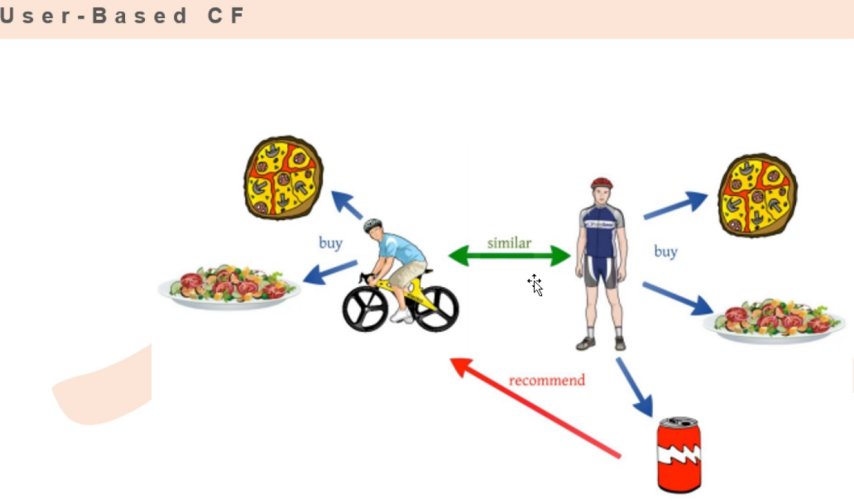




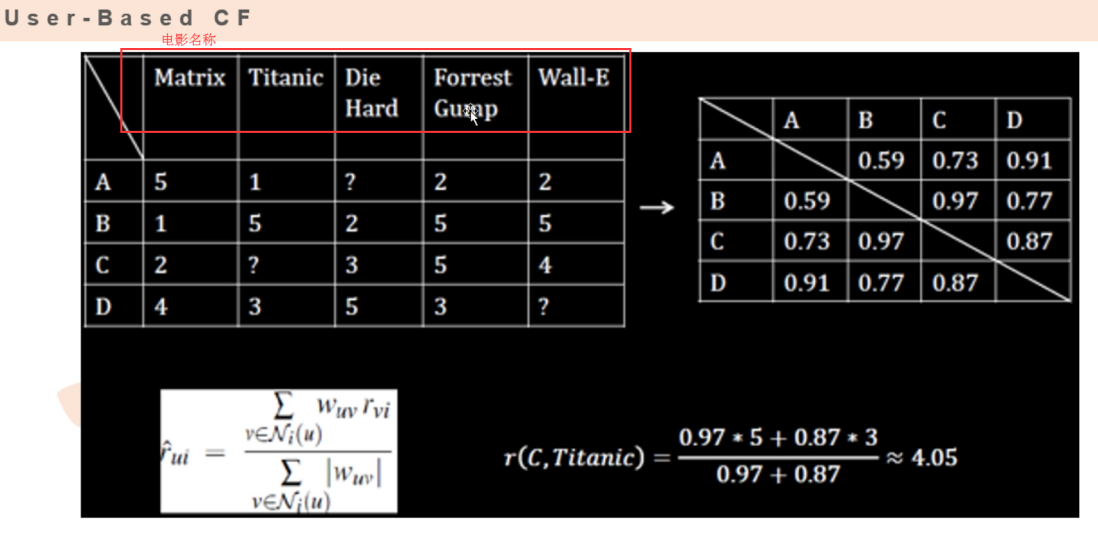


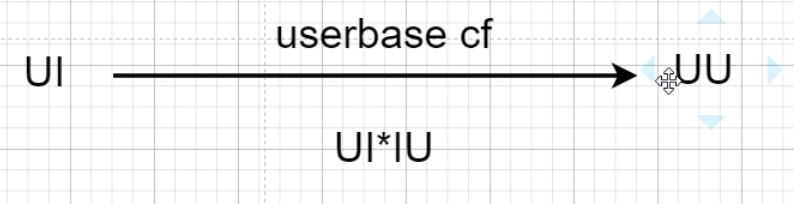
**Usercf:用户角度 itemcf:物品角度**



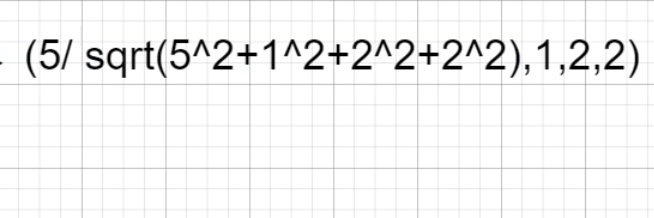
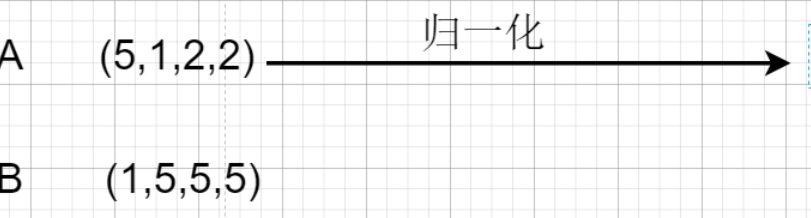


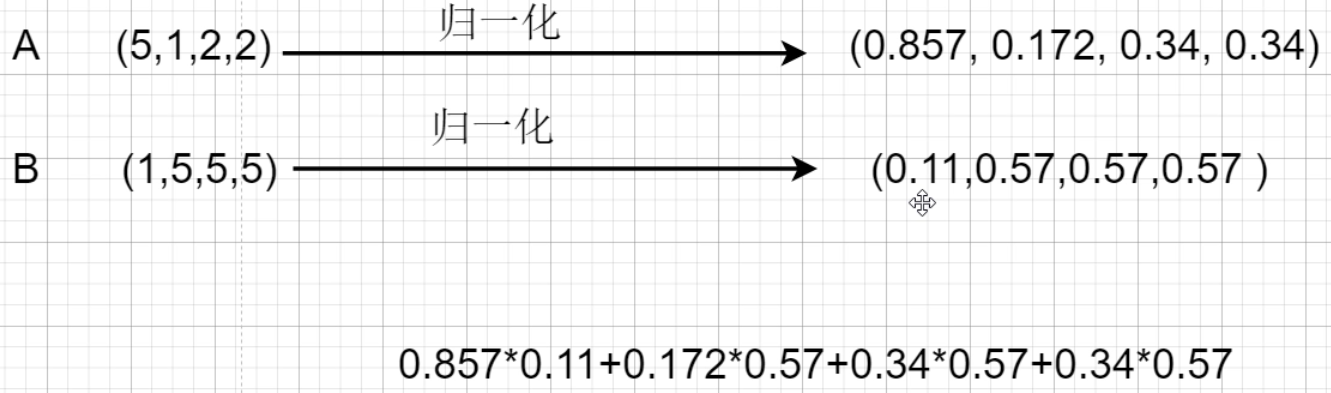
问号的是没看过的电影，求问号的电影是否可以推荐





**问号不要了**





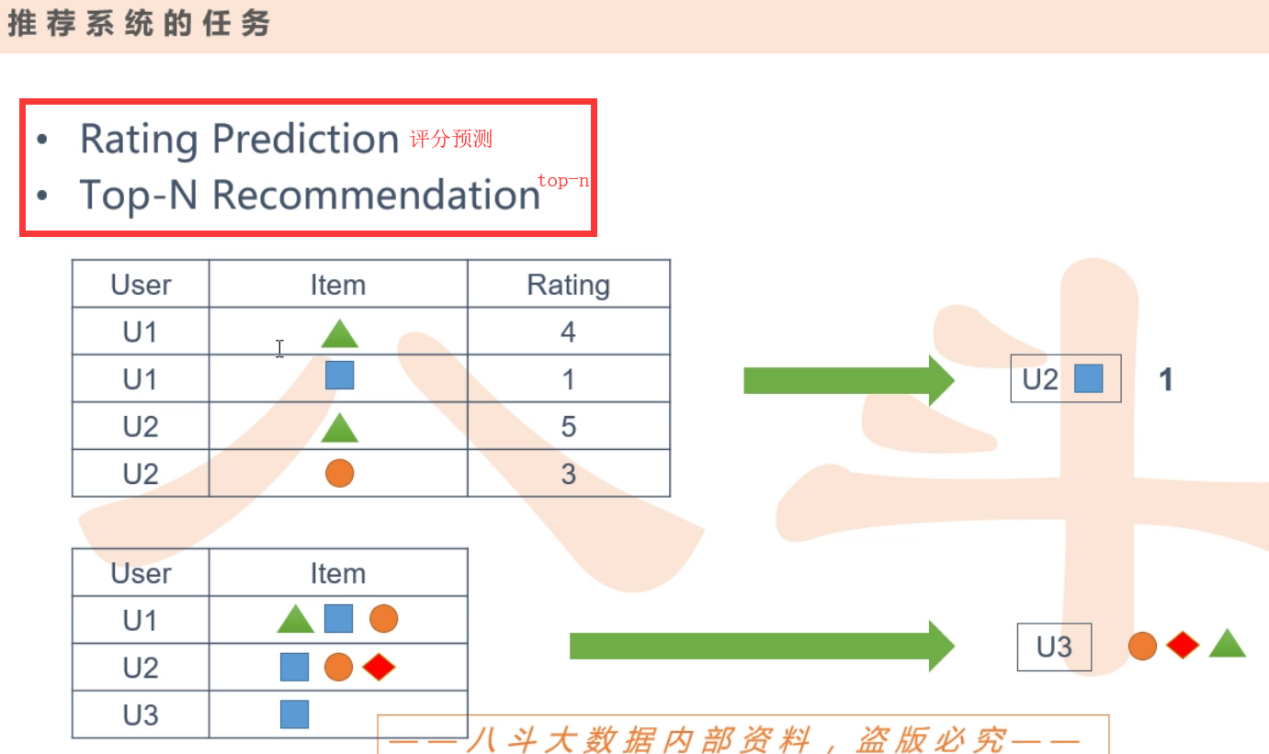
**归一化方法：**

**也是单位化**

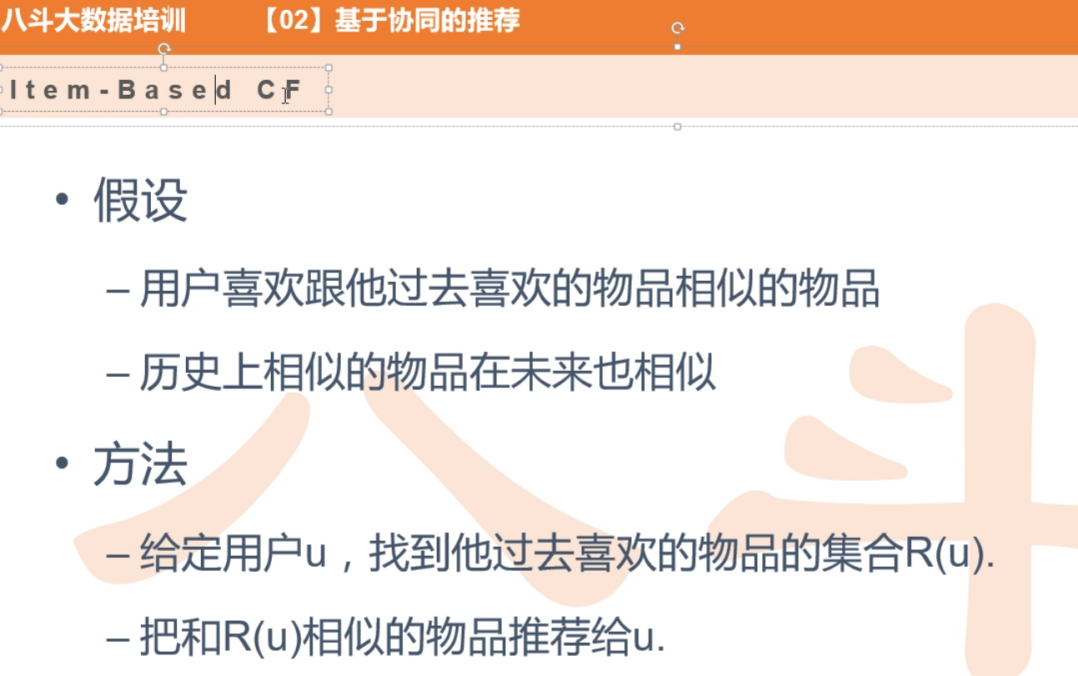
https://baike.baidu.com/item/%E5%BD%92%E4%B8%80%E5%8C%96%E6%96%B9%E6%B3%95/10089118  
1、把数变为（0，1）之间的小数  
主要是为了数据处理方便提出来的，把数据映射到0～1范围之内处理，更加便捷快速。  
2、把有量纲表达式变为无量纲表达式  
归一化是一种简化计算的方式，即将有量纲的表达式，经过变换，化为无量纲的表达式，成为纯量。

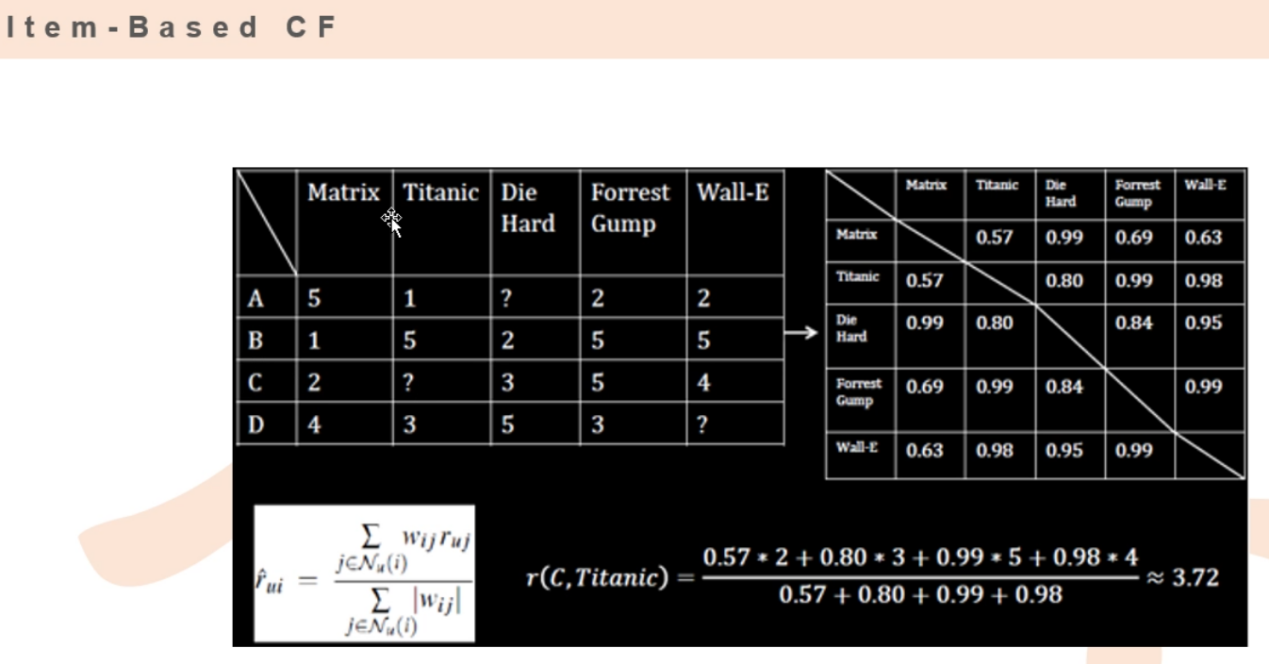
矩阵归一化：<https://blog.csdn.net/u010381985/article/details/59490102>

矩阵归一化，说白了就是整体地乘一个系数，使矩阵的绝对值=1。

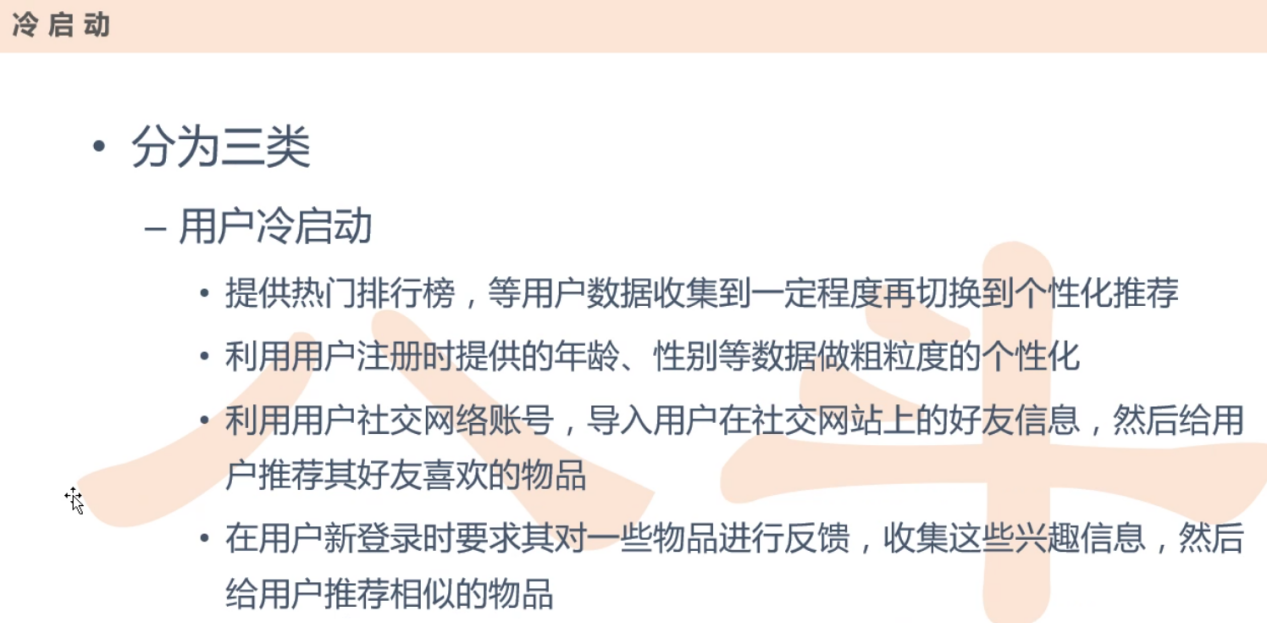


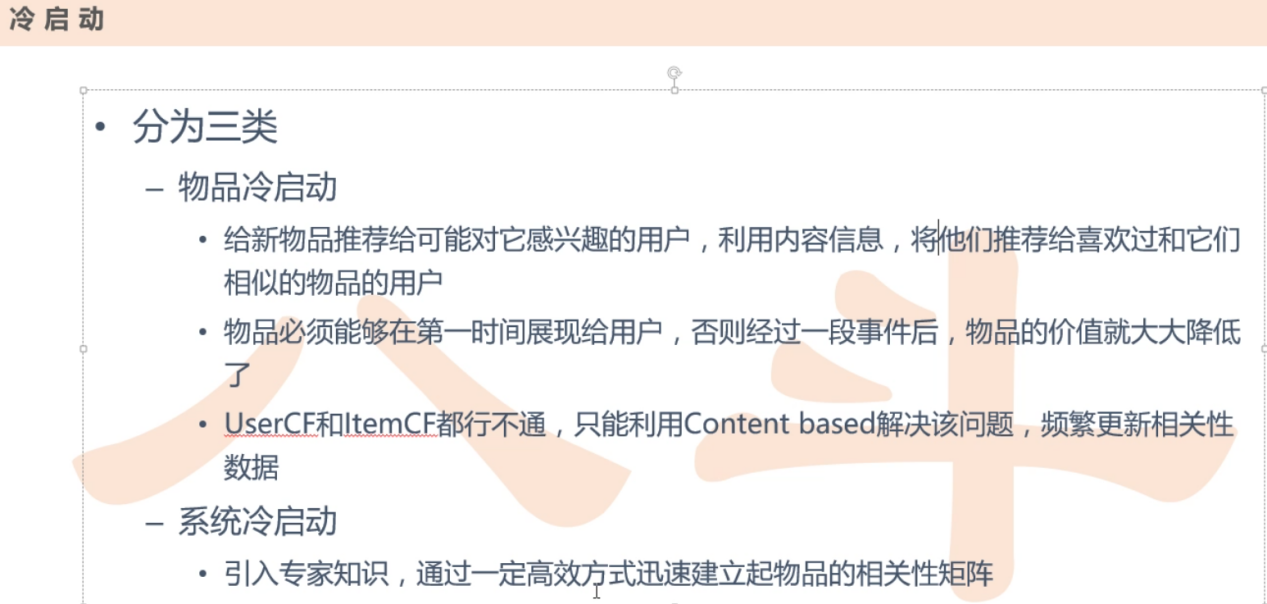
**评分预测、topn是录播讲的**





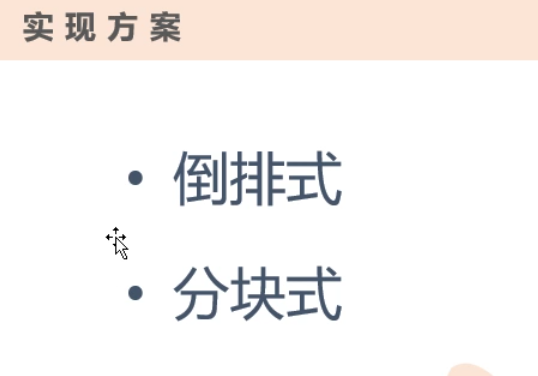




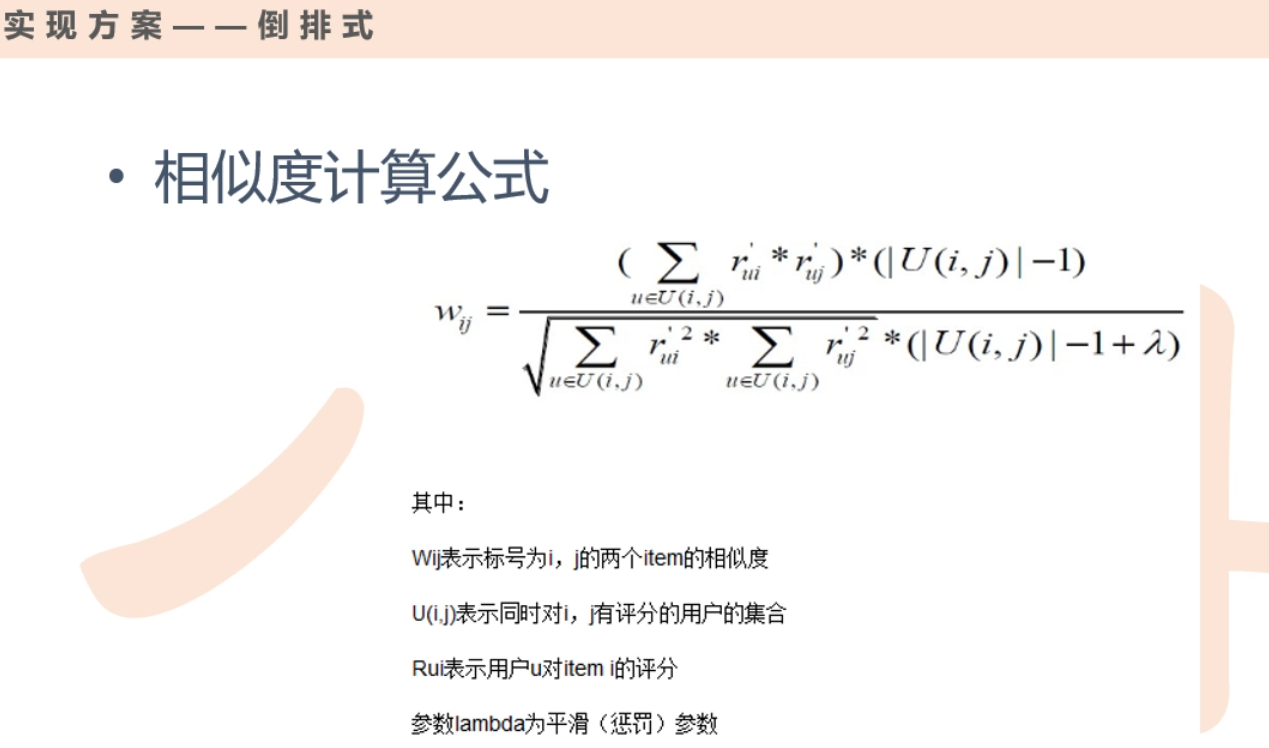




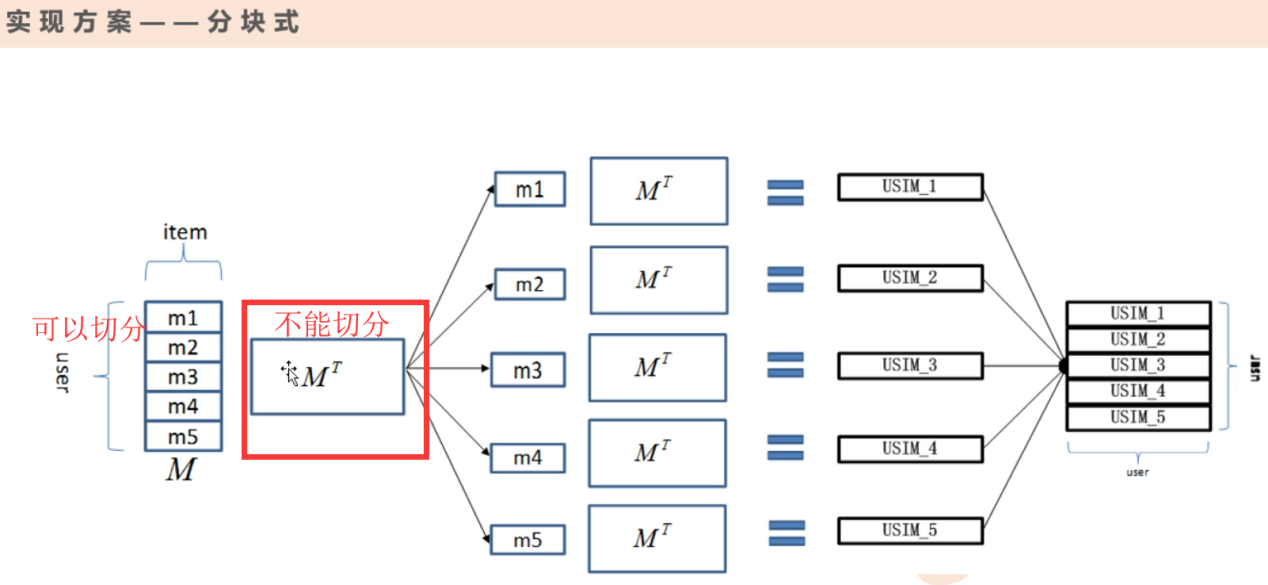
协同的实现方法









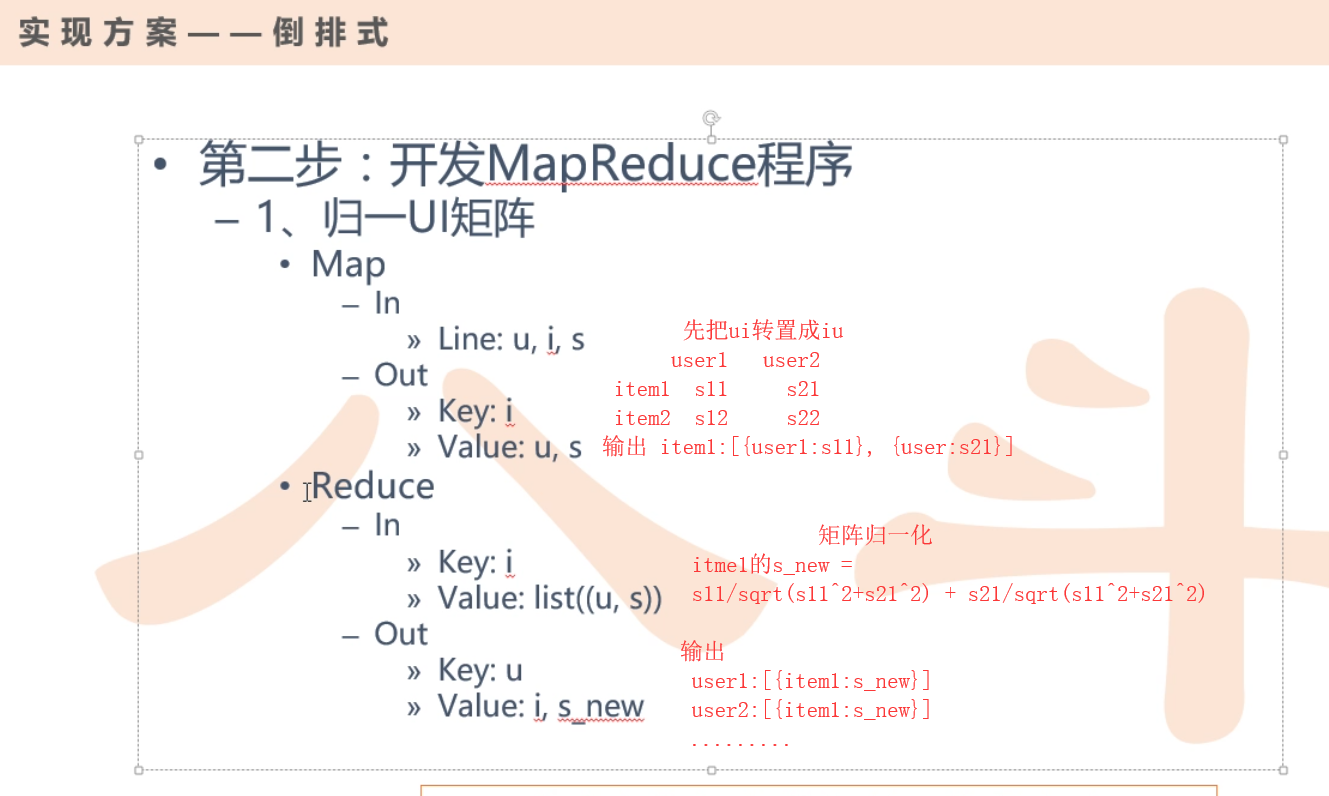


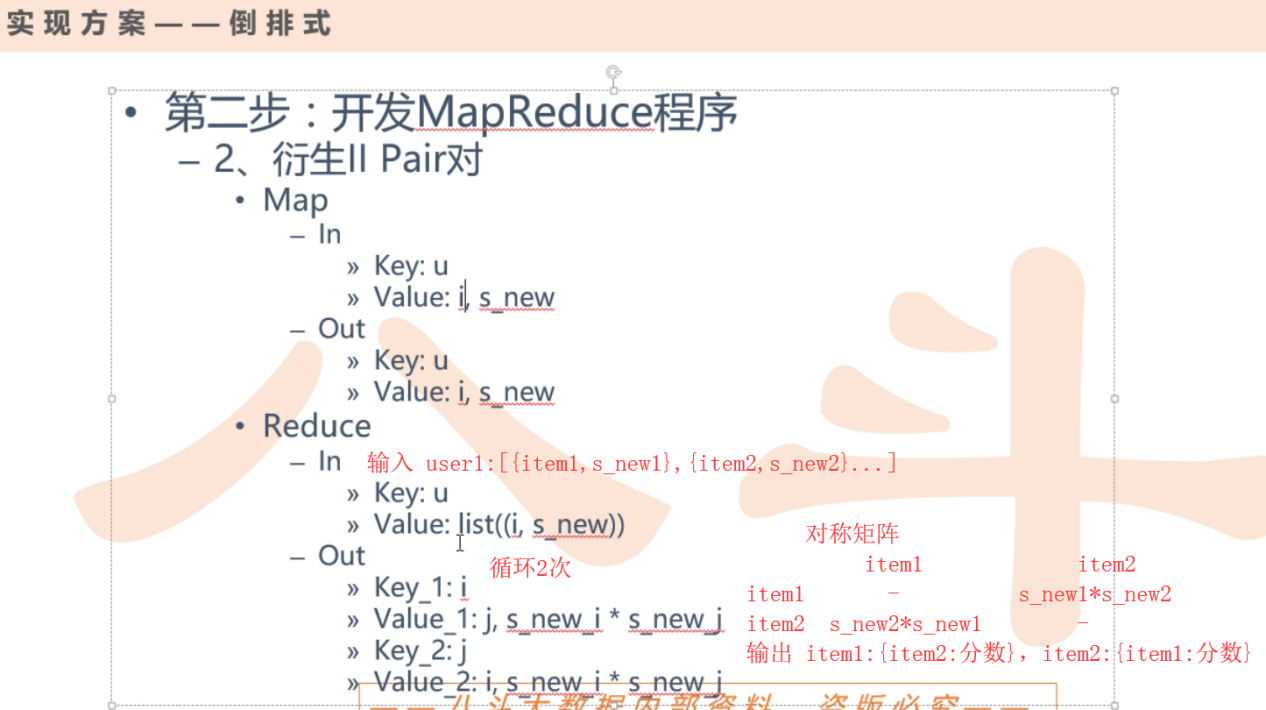
**倒排缺点：**

**如果太多数据，内存不够，桶会爆了。数据倾斜。只能截断，比如1000里面取100个。**

所以用倒排更多一点









**归一化后计算方便**

# 协同过滤和基于内容推荐有什么区别:

数据源不一样

https://www.zhihu.com/question/19971859