协同过滤算法

协同过滤算法就是依首先从字面上理解，“协同”需要一个“集体”，“过滤”就应该是晒选的意思，那么协同过滤总的来说就是通过“集体”来“筛选”，以评分推荐系统为例子，这里的“协同”我个人理解就是集合”众多人的评价”，这里的“评价”，就是“对集体都接触过的事物进行打分”，这样大概就能通过一些共同的事物反应出用户不同的”价值观“，然后通过这样的价值观来”筛选“出价值观高度相似的人，再相互推荐共同都喜爱的东西。那么这样的推荐就很有可能是大家都需要的

普及的比较多的前者，它基于关注的目标，又分为基于用户的协同过滤和基于项目的协同过滤，上面举的一个简单的评分推荐系统的例子就可以说是基于用户的协同过滤，它是通过用户对共同物品的“主观价值”来筛选相似用户，再互补评分高的商品，从而达到推荐商品的目的；那么基于项目的意思就是通过这个用户集体对商品集的评价，在物品的角度上去寻找相似度高的物品，达到推荐商品的效果。虽然针对的目标不通，但以我个人理解，大体上都是依赖这个用户集营造的“价值观”，只不过区别在于，基于用户的CF是“关心”各个用户的“主观价值”上的“区别”，而基于项目的CF则是要基于这整个用户集对项目集的“普世价值观”，来甄别出“物品”上的差异。不知道这么比喻恰不恰当哈，“普世”我这边理解就是“大多数”，是一种整体趋势的意思。价值观比较“抽象”的话，再直接点这里的“价值观”就相当于物理中的“参考系

据用户与用户之间的相似度或者物品与物品之间的相似度来进行推荐的算法，也就是依靠相似度来进行`User-Item`矩阵的填充。

一、基于记忆的协同过滤算法

1.分类

基于记忆的协同过滤算法直接对整个`User-Item`评分矩阵进行计算，通过相似度计算寻找相似近邻来产生推荐结果。主要分为以下两种：

- (1) User-Based-CF：基于用户的协同过滤
- (2) Item-Based-CF：基于物品的协同过滤

2.过程

基于记忆的协同过滤算法的`过程`主要有以下几步：

- (1) 计算相似度
- (2) 选择相似近邻
- (3) 预测评分
- (4) 推荐

3.优缺点

基于记忆的协同过滤算法的优点有：

- (1) 推荐的理论解释较清晰，易于理解
- (2) 理论上推荐精度更高

缺点有：

- (1) 维度爆炸问题 超高维度空间寻找最近邻计算代价太大

- (2) 评分矩阵过于稀疏导致最近邻准确度降低
- (3) 冷启动问题 新用户无法获得精确的推荐值

二、基于模型的协同过滤算法

基于模型的协同过滤算法会首先离线处理原始数据矩阵，得到抽象化的特征模型，从而减少高维稀疏矩阵的计算时间。

1.分类与过程

- 矩阵因子分解（降维）
  - (1) SVD 矩阵奇异值分解
    - 过程：设`User-Item`矩阵为$D{*m*n}，使用，使用*D*{m*n} = *U_{*m*i} *S_{*i*i} *V^T_{*i*n}即可得到，其中即可得到，其中i为非0奇异值的个数，而为非0奇异值的个数，而V^T_{i*n}$矩阵主要反映了物品信息，由此矩阵可以得出相似度进而进行推荐
  - (2) 交替最小二乘矩阵分解
    - 过程：交替的固定用户向量和物品向量，不断迭代来更新用户特征向量和物品特征向量，从而计算出用户对物品的预测评分进行推荐
  - (3) PCA 主成分分析
- 概率模型
  - 贝叶斯网络
- 聚类
  - K-Means聚类
  - 模糊K-Means聚类
- 关联规则

2.优缺点

基于模型的协同过滤算法的优点有：

- 速度较快，由于建立的模型维度通常比原数据集小，因此速度较快

缺点有：

- 不够直观，通常基于模型的CF算法的理论解释要难于基于记忆的CF算法