推荐算法(1)：协同过滤总结

推荐算法(2)：基于内容的推荐

推荐算法（3）：利用用户标签数据

推荐算法（4）利用上下文信息

推荐算法（5）利用社交网络数据

推荐算法（6） 实例

推荐算法(7)缺失的评分预测问题

推荐算法（8）评测指标

冷处理

一、协同过滤方法：

（1）基于内容/基于领域的协同过滤

ItemCF：计算items之间的相似度，推荐与A的已知item最相关的item

步骤：

1.输入item-user矩阵

2.求item-item相似度 （不同相似度度量：person系数、曼哈顿距离、余弦值、欧几里得距离）

3.根据相似度求评分

UserCF：计算user之间的相似度，推荐相似用户A的item给B。

步骤：

1.输入item-user矩阵

2.求user-user相似度（不同相似度度量：person系数、曼哈顿距离、余弦值、欧几里得距离）

3.根据相似度求评分

（2）基于Model的协同过滤

1.用关联算法做协同过滤

（1）Apriori

算法流程：

下面我们对Aprior算法流程做一个总结。

输入：数据集合D，支持度阈值α

　　　　输出：最大的频繁k项集

1）扫描整个数据集，得到所有出现过的数据，作为候选频繁1项集。k=1，频繁0项集为空集。

2）挖掘频繁k项集

a) 扫描数据计算候选频繁k项集的支持度

b) 去除候选频繁k项集中支持度低于阈值的数据集,得到频繁k项集。如果得到的频繁k项集为空，则直接返回频繁k-1项集的集合作为算法结果，算法结束。如果得到的频繁k项集只有一项，则直接返回频繁k项集的集合作为算法结果，算法结束。

c) 基于频繁k项集，连接生成候选频繁k+1项集。

3） 令k=k+1，转入步骤2。

（2）FPGrowth

FP Tree算法包括三步：

1）扫描数据，得到所有频繁一项集的的计数。然后删除支持度低于阈值的项，将1项频繁集放入项头表，并按照支持度降序排列。

2）扫描数据，将读到的原始数据剔除非频繁1项集，并按照支持度降序排列。

3）读入排序后的数据集，插入FP树，插入时按照排序后的顺序，插入FP树中，排序靠前的节点是祖先节点，而靠后的是子孙节点。如果有共用的祖先，则对应的公用祖先节点计数加1。插入后，如果有新节点出现，则项头表对应的节点会通过节点链表链接上新节点。直到所有的数据都插入到FP树后，FP树的建立完成。

4）从项头表的底部项依次向上找到项头表项对应的条件模式基。从条件模式基递归挖掘得到项头表项项的频繁项集。

5）如果不限制频繁项集的项数，则返回步骤4所有的频繁项集，否则只返回满足项数要求的频繁项集。

2.用聚类

(1)kmeans

(2)minibatchkmeans

分批数据进去训练

(3)birch 层次聚类

数据集 建立CF tree

再在CF tree的基础上聚类

3.用分类

（1）贝叶斯分类器

某个人对物品的评分的概率，可以用贝叶斯的后验概率求出

（2）逻辑回归

（3）KNN

4.用回归

5.隐语义

（1）用矩阵分解：

基于SVD推荐

先分解成三个矩阵，再新构成两个相关矩阵，最后进行评分

(2)Funk SVD (LFM) （基于NMF推荐）

在SVD的基础上，加上了正则项，对p,q进行限制。

(3)LFM改进 (加偏置)

利用隐语义模型可以对评分矩阵中的缺省值进行预测，但是模型没有考虑用户和物品本身的属性，现实中用户和物品属性本身有时就会存在较大的差异。例如乐观用户对物品的评分要普遍高于消极用户对物品的评分，质量好的物品获得的评价要普遍高于质量劣的物品。即使他们对同一物品的评分相同，他们对该物品的偏好程度也可能不同，因此需要在原有的模型中加入偏置因子来区别不同的用户和物品。

u表示全局平均数

bu表示表示用户评分相对于平均值的偏移量，不受物品属性的影响。对于乐观用户ub会偏高，消极用户ub会偏低

bi 表示物品偏执，对于优质物品,bi高，低质的bi低。

(4)SVD++（加入隐式信息)

SVD算法是指在SVD的基础上引入隐式反馈，使用用户的历史浏览数据、用户历史评分数据、电影的历史浏览数据、电影的历史评分数据等作为新的参数。

（5）TrustSVD （加入社会化信任机制）

现实中朋友彼此之间会受对方的影响，朋友推荐的物品更具有说服力和信任度。

（6）改进的信任模型

不同人信任程度不一样

基本原理是如果一个用户被越多的人信任，他的可信度就会越高，别人对他的信任权值就会越高

6.用神经网络

(1)wide&deep

7.用图模型

（1）基于随机游走的PersonalRank算法

（2） 改进 用矩阵求解的方式

二、对于协同过滤的评分方法：

偏好收集

显式：评分、投票、转发、保存书签、标记标签、评论

隐式：点击、页面停留、购买

数据稀疏：

1.基于图的方法 ： 假定用户品味的传递性，并由此增强额外信息矩阵

可以提高推荐质量，尤其在矩阵稀疏的时候

2.缺省投票：给那些一两个用户评过分的物品赋以缺省值

冷启动：

可以用到用户的附加信息（年龄，性别，教育程度等）

————————————————

版权声明：本文为CSDN博主「礼拜天吃芋圆」的原创文章，遵循CC 4.0 BY-SA版权协议，转载请附上原文出处链接及本声明。

原文链接：https://blog.csdn.net/weixin\_38636668/article/details/90297133