1. 推荐系统技术概览

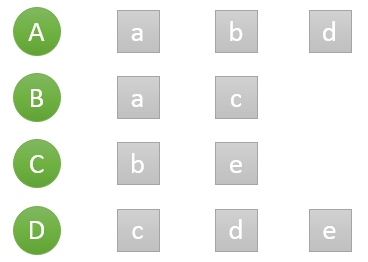
深度学习推荐系统之前

1. 基于用户的协同过滤
2. 原理

构建用户为行坐标，商品为列坐标的共现矩阵，在生成共现矩阵之后，推荐问题就转换为预测矩阵中问号元素的问题（其实是计算用户与用户向量之间的余弦相似度）。

1. 问题

用户数往往大于物品数，所以存储共现矩阵的开销很大；同时用户的数据向量往往会十分稀疏，导致找到相似用户的准确度是非常低的。



1. 基于物品的协同过滤
2. 原理

通过计算共现矩阵中物品列向量的相似度得到物品之间的相似矩阵，因为一个用户购买过物品a和物品b，说明a，b之间有相似性。再找到用户的历史正反馈物品的相似物品进行进一步排序和推荐

1. 问题

共现矩阵的稀疏性，出现推荐结果的头部效应明显。

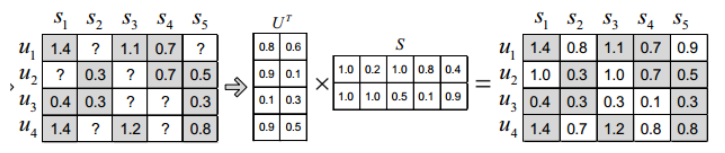


1. 矩阵分解算法 --- 协同过滤的进化，使任意的用户和物品之间都可以得到预测分值。
2. 原理

矩阵分解算法则期望为每一个用户和视频生成一个隐向量，将用户和视频定位到隐向量的表示空间上，距离相近的用户和视频表明兴趣特点接近。在推荐过程中就应该把距离相近的视频推荐给目标用户。矩阵分解算法将m \* n维的共现矩阵分解为m \* k维的用户矩阵U，和k \* n维的物品矩阵V。其中m为用户的数量，k为隐向量的维度。K越小，隐向量的表达能力弱但是泛化能力强。

1. 存在的问题

没有办法加入用户和物品上下文的信息。同时在缺乏用户历史行为时，无法进行有效的推荐。



1. 逻辑回归 --- 融合多种特征的推荐模型
2. 原理

极大似然法更新模型参数，在考虑融合不同特征，形成比较全面的推荐结果。同时逻辑回归的数学意义满足CTR点击事件的实际意义。（点击还是不点击）。

1. 存在的问题

无法进行特征交叉、特征筛选等一系列高级操作。在仅使用单一特征而不是交叉特征进行判断的情况下，有时不仅是信息损失的问题，甚至会得出错误的结论。

从FM到FFM --- 自动特征交叉的解决方案

1. POLY2模型 --- 对特征进行暴力组合
2. 原理

该模型对所有特征进行两两交叉，并对所有的特征组合赋予权重。但POLY2模型本质上仍是线性模型，其训练方法与逻辑回归并无区别

1. 存在的问题

参数的增加，加大了训练的复杂度；该方法使得原本稀疏的向量更为稀疏。

1. FM模型 --- 隐向量特征交叉
2. 原理

隐向量的引入可以很好的解决数据稀疏的问题。在某推荐场景下，样本有两个特征，分别是频道（channel）、品牌（brand），某训练样本的特征组合是（ESPN，Adidas）。在POLY2中只有当ESPN和Adidas都出现时，模型才能学习到这个组合特征的权重，而在FM中ESPN的隐向量也可以通过（ESPN,Gucci）样本进行更新，Adidas的隐向量也可以通过（NBC，Adidas）样本进行更新，这大幅降低了模型对数据稀疏性的要去。设置一个从未出现的特征组合也可以具备隐向量。FM提高了泛化能力。

1. 存在的问题
2. FFM模型 --- 引入特征域的概念
3. 原理
4. 存在的问题
5. GBDT + LR --- 特征工程模型化的开端
6. 原理

利用GBDT（决策树组成的森林）进行特征的选取

1. 存在的问题
2. 存在的问题
3. 位置偏置

用户点击某个广告，并非出于喜好，有可能仅仅与展示位置有关。

1. 选择偏置

物品只在相对靠后的位置、展示给了少量的用户，属于尾部物品。也就是说，选择偏置是推荐系统对长尾物品推荐准确性不高的一个原因。

1. 马太效应

强者恒强，弱者愈弱