

1. 算法选择：首先查看用户物品交互数据，训练集包含了用户和物品的点击数据，是隐式交互数据而非显式交互数据，并且作业的目的是为每个用户做 top10 推荐，于是我想到了课上讲过的 BPR (Bayesian Personalized Ranking)。由于隐式反馈数据有诸多弊端，比如隐式反馈数据不明确、具有噪声等等，在使用隐式反馈数据的情况下，用户与物品之间的点击数据被视为正例，因为只有用户和物品产生交互之后我们才能收集到数据；而对于哪些没有被观察到的数据，即用户和物品之间没有产生交互数据，则分为两种情况，一种是用用户对某些物品确实没有兴趣，另一种是缺失值，即用户以后可能会与该物品产生交互行为，在传统的个性化推荐中，通常都是将所有观察到的隐式数据视为正类，而其余的视为负类。

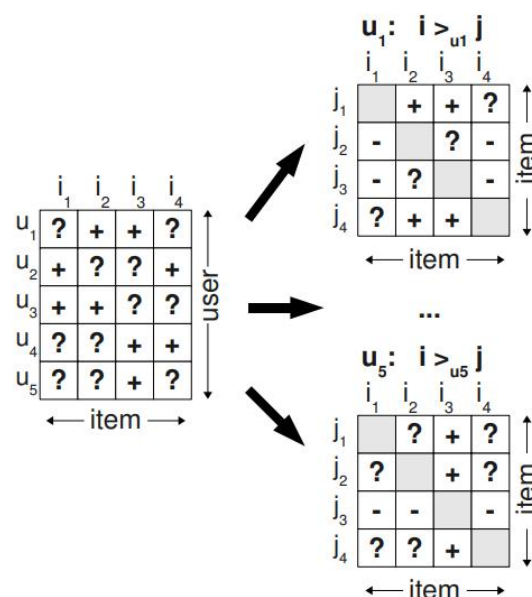
贝叶斯个性化推荐算法与传统推荐模型不同的是，BPR 通过对问题进行贝叶斯分析后得到的最大后验概率来对物品进行排序，从而产生个性化推荐。

2. 模型实现思路：查阅论文 [BPR: Bayesian Personalized Ranking from Implicit Feedback](#)，BPR 首先将用户与物品的隐式反馈数据转化为三元组  $(u, i, j)$ ，它代表用户  $u$  对于物品  $j$  来说更喜欢物品  $i$ ，我们假设用户  $u$  对交互过的物品的喜爱程度大于他没有交互过的物品的喜爱程度，故三元组  $(u, i, j)$  即用户  $u$  交互过物品  $i$  而未交互过物品  $j$ ，即：

$$D_S := \{(u, i, j) | i \in I_u^+ \wedge j \in I \setminus I_u^+\}$$

其中  $I$  是总物品集合， $I_u^+$  是用户  $u$  交互过的物品集合。

所以对于每个用户都可以构建一个偏好矩阵：



对于三元组  $(u, i, j)$ ， $i$  是产生过交互的物品， $j$  是未产生交互的物品，所以构建的三元组只包含图中标签为+的数据。根据 BPR 算法可以将代码的思路大致划分为：数据初始化 -> 初始化参数，包括学习率、正则化参数等 -> 训练模型 -> 评估模型 -> 生成推荐列表并写入文件。

BPR 算法优化部分使用了随机梯度下降法，具体的优化函数和函数对超参的偏导推导过程详情在论文中，其中算法的优化过程如下图所示：

```

1: procedure LEARNBPR( $D_S, \Theta$ )
2:   initialize  $\Theta$ 
3:   repeat
4:     draw  $(u, i, j)$  from  $D_S$ 
5:      $\Theta \leftarrow \Theta + \alpha \left( \frac{e^{-\hat{x}_{uij}}}{1+e^{-\hat{x}_{uij}}} \cdot \frac{\partial}{\partial \Theta} \hat{x}_{uij} + \lambda_{\Theta} \cdot \Theta \right)$ 
6:   until convergence
7:   return  $\hat{\Theta}$ 
8: end procedure

```

其中 $x_{uij}$ 表示用户  $u$  对物品  $i$  的评分与用户  $u$  对物品  $j$  的评分之差,即 $x_{uij} = x_{ui} - x_{uj}$ , $\lambda_{\theta}$ 表示正则化参数,  $\theta$ 表示超参数。

优化函数对各个超参数的偏导部分如下图所示:

$$\frac{\partial}{\partial \theta} \hat{x}_{uij} = \begin{cases} (h_{if} - h_{jf}) & \text{if } \theta = w_{uf}, \\ w_{uf} & \text{if } \theta = h_{if}, \\ -w_{uf} & \text{if } \theta = h_{jf}, \\ 0 & \text{else} \end{cases}$$

其中 $w_{uf}$ 表示用户  $u$  隐向量中的第  $f$  个元素; $h_{if}$ 表示物品  $i$  隐向量中的第  $f$  个元素, $h_{jf}$ 表示物品  $j$  隐向量中的第  $f$  个元素。对于其它参数, 偏导为 0. 这段梯度部分具体体现在函数`train\_bpr`中, BPR 旨在通过最大化用户实际喜欢的物品的预测评分与不喜欢的物品的预测评分之差来优化用户个性化推荐。