

1-7

今天主要对以后的工作做个规划。还没有想到创新点，个人感觉还是在以后的工作中想创新点吧。

参考其他论文计划使用的baseline：

- 贝叶斯个性化排序模型（BPR）
- 普通的矩阵分解模型（MF）
- 基于倾向分数的矩阵分解模型(MF-IPS)
- 相关性矩阵分解模型（RMF）

这些是比较经典的模型，不计划把sota作为baseline，没有信心做出比sota更好的效果。

BPR学习

BPR算法的几个核心点：

- 每个用户之间的偏好行为相互独立
- 同一用户对不同物品的偏序相互独立
- $\langle u, i, j \rangle$ 表示用户 u 对物品 i 的喜好大于物品 j ，写成 $i >_u j$
- 使用最大后验估计计算参数

\bar{X} 表示用户集 U 和物品集 I 对应的 $U \times I$ 的预测排序矩阵，分解后的用户矩阵为 W 和物品矩阵为 H 。

对于任意一个用户 u ，对应任意一个物品 i 有：

$$\bar{x}_{ui} = \omega_u \bullet h_i = \sum_{f=1}^k \omega_{uf} h_{if}$$

最终目标是找到合适的 W 和 H ，让 \bar{X} 和 X 最相似。

BPR的优化思路：

BPR是基于最大后验估计 $P(W, H | >_u)$ 来求解模型参数 W 和 H ，这里用 θ 来表示 W 、 H ，用 $>_u$ 来代表用户 u 对应的所有物品的全序关系，转为优化 $P(\theta | >_u)$ 。

$$P(\theta | >_u) = \frac{P(>_u | \theta) P(\theta)}{P(>_u)}$$

对任意用户 u 来说， $P(>_u)$ 对所有物品都一样，所以有：

$$P(\theta | >_u) \propto P(>_u | \theta) P(\theta)$$

第一部分 $P(>_u | \theta)$ 和数据集有关，第二部分 $P(\theta)$ 和数据集无关。

然后每个用户之间的偏好行为相互独立，同一用户对不同物品偏序相互独立，所以有：

$$\prod_{u \in U} P(>_u | \theta) = \prod_{(u,i,j) \in (U \times I \times I)} P(i >_u j | \theta)^{\delta((u,i,j) \in D)} (1 - P(i >_u j | \theta))^{\delta((u,i,j) \notin D)}$$

其中

$$\delta(b) = \begin{cases} 1 & \text{if } b \text{ is true} \\ 0 & \text{else} \end{cases}$$

第一部分可以简化为：

$$\prod_{u \in U} P(>_u | \theta) = \prod_{(u,i,j) \in D} P(i >_u j | \theta)$$

对于 $P(i >_u j | \theta)$ 这个概率，使用下面这个式子代替：

$$P(i >_u j | \theta) = \sigma(\bar{x}_{uij}(\theta))$$

$\sigma(x)$ 是sigmoid函数，在此使用为了方便优化计算。

对于 $\bar{x}_{uij}(\theta)$ 这个式子，要满足当 $i >_u j$ 时， $\bar{x}_{uij}(\theta) > 0$ ，当 $j >_u i$ 时， $\bar{x}_{uij}(\theta) < 0$ ，

所以可以另 $\bar{x}_{uij} = \bar{x}_{ui} - \bar{x}_{uj}$ ，

对于 $\bar{x}_{ui}(\theta)$, $\bar{x}_{uj}(\theta)$ ，就是矩阵 \bar{X} 对应位置的值。

最终，第一部分的优化目标转化为：

$$\prod_{u \in U} P(>_u | \theta) = \prod_{(u,i,j) \in D} \sigma(\bar{x}_{ui} - \bar{x}_{uj})$$

对于第二部分，直接假设他服从均值为0，协方差矩阵为 $\lambda_\theta I$ 的高斯分布：

$$P(\theta) \sim N(0, \lambda_\theta I)$$

然后最大后验估计函数得知，使用梯度下降法求解参数。

指标相关：

- 准确度指标：平均绝对误差（MAE）、均方误差根（RMSE）
- 集合推荐指标，用于Top-N推荐任务：精密度(Precision)、召回(Recall)、AUC和NDCG@K

后续工作：

把指标全都实现，留出统一接口，将来直接使用即可。

baseline也是，这些固定的东西要先完成，创新点的提出感觉是在这个过程中产生的。