八、推荐算法

一、推荐系统的评估

（1）打分

（2）TOP-N 准确度、召回率

□ 准确度：

❷ Top N推荐

设R(u)为根据训练建立的模型在测试集上的推荐，
T(u)为测试集上用户的选择。

$$Precision = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|}$$

$$Recall = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|}$$

（3）覆盖率，尽可能推荐全部商品，熵最大

（4）多样性，对于同一个人推荐不同样别的（推荐样品中，两两之间不相似度）

二、基于内容的推荐

□ 基于内容的推荐

- 对于每个要推荐的内容，我们需要建立一份资料：
  - 比如词k_j在文件d_j中的权重w_{ij}
  - 常用的方法比如TF-IDF
- 需要对用户也建立一份资料：
  - 比如说定义一个权重向量$(w_{c1}, \cdots, w_{ck})$
  - 其中w_{ci}表示第ki个词对用户c的重要度
- 计算匹配度
  - 比如用余弦距离公式

$$u(c,s) = \cos(\vec{w}_c, \vec{w}_s) = \frac{\vec{w}_c \cdot \vec{w}_s}{\|\vec{w}_c\|_2 \times \|\vec{w}_s\|_2} = \frac{\sum_{i=1}^{K} w_{i,c} w_{i,s}}{\sqrt{\sum_{i=1}^{K} w_{i,c}^2} \sqrt{\sum_{i=1}^{K} w_{i,s}^2}}$$

三、协同过滤

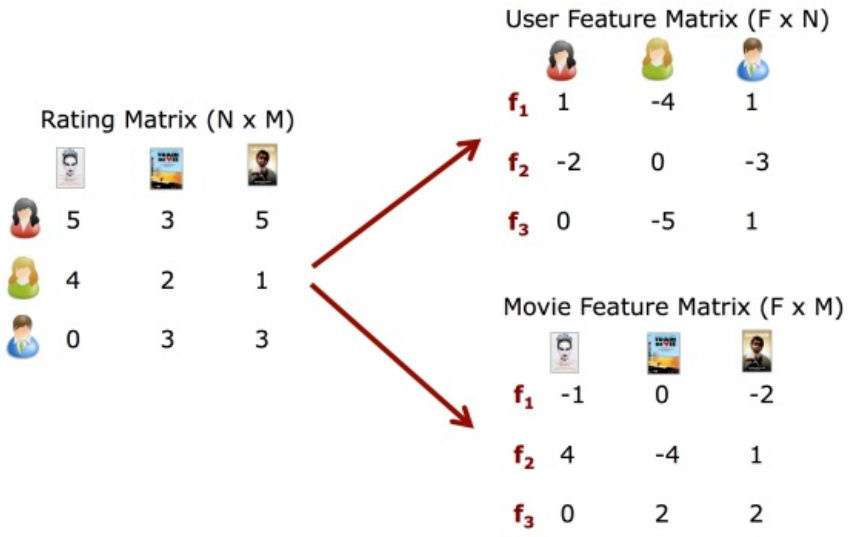
基于Item的，基于User的，要注意评分矩阵减去均值，比如某个用户给所有打分较高，另一个用户对所有用户打分较低，个人标准不同

四、隐语义模型

隐喻义模型假设矩阵R可以分解成P和Q，然后PQ乘回去就可以填充协同矩阵中R为0 的值

解释原因：

隐语义模型



隐喻义模型的求解，平方损失函数，梯度下降法

矩阵分解

- 假定有U个用户，D个item，R为打分矩阵
- 假定有K个隐含变量，我们需要找到矩阵P(U*K)和Q(D*K)：

$$\mathbf{R} \approx \mathbf{P} \times \mathbf{Q}^T = \hat{\mathbf{R}}$$

$$\hat{r}_{ij} = p_i^T q_j = \sum_{k=1}^{k} p_{ik} q_{kj}$$

- 如何才能找到最佳的P和Q呢？
  - 梯度下降
    1. 定义损失函数

$$e_{ij}^2 = (r_{ij} - \hat{r}_{ij})^2 = (r_{ij} - \sum_{k=1}^{K} p_{ik} q_{kj})^2$$

    2. 求解梯度

$$\frac{\partial}{\partial p_{ik}} e_{ij}^2 = -2(r_{ij} - \hat{r}_{ij})(q_{kj}) = -2e_{ij} q_{kj}$$
$$\frac{\partial}{\partial q_{ik}} e_{ij}^2 = -2(r_{ij} - \hat{r}_{ij})(p_{ik}) = -2e_{ij} p_{ik}$$

优化，有可能造成过分符合R， 乘回去可能0的仍然为0，加入正则项

矩阵分解

- 别忘了正则化：

$$e_{ij}^2 = (r_{ij} - \sum_{k=1}^{K} p_{ik} q_{kj})^2 + \frac{\beta}{2} \sum_{k=1}^{K} (||P||^2 + ||Q||^2)$$

- 再次求梯度/偏导，更新迭代公式：

$$p'_{ik} = p_{ik} + \alpha \frac{\partial}{\partial p_{ik}} e_{ij}^2 = p_{ik} + \alpha(2e_{ij} q_{kj} - \beta p_{ik})$$
$$q'_{kj} = q_{kj} + \alpha \frac{\partial}{\partial q_{kj}} e_{ij}^2 = q_{kj} + \alpha(2e_{ij} p_{ik} - \beta q_{kj})$$

- 再还原回矩阵乘积，即可补充未打分项
- 通常情况下，我们会限定分解得到的P和Q中的元素都非负，这样得到的结果是一定程度上可解释的。
- 因为不存在减法操作，因此可以看做对隐变量特征的线性加权拟合。

隐语义模型的进一步优化，假设每个打分项受电影bias，用户bias，以及所有评分均值和PQ的双重影响

$$r_{xi} = \underbrace{\mu}_{\text{Overall mean rating}} + \underbrace{b_x}_{\text{Bias for user } x} + \underbrace{b_i}_{\text{Bias for movie } i} + \underbrace{q_i \cdot p_x}_{\text{User-Movie interaction}}$$

此时的损失函数为：

- 加bias的隐语义模型
  - 需要最小化

$$\min_{Q,P} \sum_{(x,i)\in R} \underbrace{\left(r_{xi} - (\mu + b_x + b_i + q_i\, p_x)\right)^2}_{\text{goodness of fit}} + \underbrace{\left(\lambda_1 \sum_i \|q_i\|^2 + \lambda_2 \sum_x \|p_x\|^2 + \lambda_3 \sum_x \|b_x\|^2 + \lambda_4 \sum_i \|b_i\|^2 \right)}_{\text{regularization}}$$

λ is selected via grid-search on a validation set