

MatrixFactorization

矩阵分解推荐算法

赵海臣

背景

▶ 推荐系统

- 基于内容的过滤(Content Filtering): 建立用户画像, 通过画像进行过滤
 - 适合解决冷启动, 通常精确度比协同过滤要差。
- 协同过滤(Collaborative Filtering): 基于用户的历史数据进行反馈过滤
 - 通常比Content-Based推荐更加精确, 但由于冷启动缺乏相应的历史数据, 所以冷启动存在问题。

协同过滤

- ▶ 协同过滤主要有两个领域：
 - 邻域模型(neighborhood method)
 - 计算物品-物品之间，用户-用户之间的关系
 - 一个物品的邻域是用户们倾向于给出相同评价的其它物品。
 - 一个用户的领域是在不同的商品上倾向于给出一致评价的其它用户。
 - 隐语义模型(latent factor method)
 - 尝试通过数学方式从用户对物品的评价中分析出物品、用户各自的特征，从而解释为何用户如此评价、为何物品获此评价

矩阵分解模型

- ▶ 矩阵分解模型能够将商品、用户都映射到一个“特征空间”(Joint Latent Factor Space), 从而可以使用空间点积的方式来描述将用户、商品的关系。
 - 每个商品被描述成特征空间里的一个向量 q_i
 - 每个用户被描述成特征空间里的一个向量 u_i
 - 用户对某个商品所表现特征的兴趣体现在二者向量的点积:
 - $r = q_i \cdot u_i$

SVD的困难

- ▶ 矩阵分解模型很类似SVD(Singular Value Decomposition)奇异值分解，但是SVD应用于协同过滤的困难在于：
 - 用户-商品(评分)交互矩阵过于稀疏
 - 标准SVD在交互矩阵不完整的情况下难以计算
 - 仅仅考虑矩阵中已有的交互数据很容易造成过拟合
- ▶ 矩阵分解模型早期的SVD解决方式填补缺失的矩阵元素，使得矩阵变稠密，但是：
 - 它将显著增加处理的数据量
 - 不精确的填补结果将严重扭曲数据

MF方程

- ▶ 通过现有的用户-商品交互数据来形成模型：
- ▶ 为了学习用户、商品的特征向量，可以通过最小化现有所有评分的预测误差error，当预测误差error达到最小值时，即获得了我们最佳预测模型。

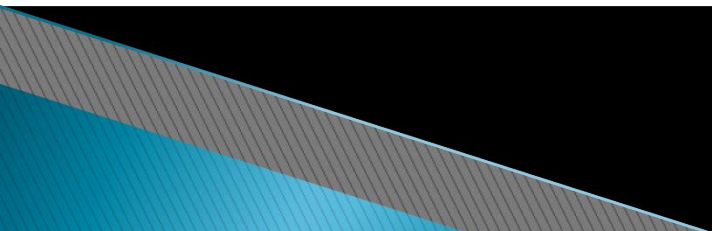
- ▶ MF方程：

$$\min_{q^*, p^*} \sum_{(u,i) \in K} (r_{ui} - q_i^T p_u)^2 + \lambda (\|q_i\|^2 + \|p_u\|^2)$$

- 其中 K 是已知的用户-商品(u,i)交互数据 r_{ui}
- 方程通过现有的观测数据来学习现有数据的规律，并用来预测缺失的数据
- 后面部分为了防止学习过程中出现过拟合， λ 用于调节步长，具体的值需要通过交叉验证来确定

MF方程的收敛算法

- ▶ MF方程的收敛算法
 - SGD(Stochastic Gradient Descent)随机梯度下降法
 - ALS(Alternating Least Squares)交替最小二乘法



SGD方法(LMS变种)

- ▶ Error误差方程:

$$e_{ui} = r_{ui} - q_i^T p_u$$

- ▶ 通过现有观测数据迭代逼近收敛 :

$$q_i = q_i + \gamma \cdot (e_{ui} \cdot p_u - \lambda \cdot q_i)$$

$$p_u = p_u + \gamma \cdot (e_{ui} \cdot q_i - \lambda \cdot p_u)$$

λ : 步长

γ : 比例参数

SGD方法(LMS变种)

▶ 计算过程：

- 假设所有商品、用户特征为 n 维描述
- 比如拟合某一个商品 i 特征列，已知对其发生行为的用户 $\{u\}$ 集合
- 对商品 i 的特征列一个随机初始值，用户 $\{u\}$ 随机初始值，假设 $\{u\}$ 的特征是确定的，并且有实际观测的 $\{r_{ui}\}$ ，所以可以对用户一个接一个地迭代
 - 迭代，直到error稳定低于某个设定阈值
 - 计算 u 对 i 的残差error
 - 将残差error反馈修正商品的特征列

ALS方法

- ▶ 如果我们将用户、商品中的任意一个矩阵修复上，那么就能够获得另外一个矩阵的最优解
- ▶ ALS的思想在于来回地迭代修复用户、商品矩阵，每次假设其中一个矩阵是最优的，从而来修复另外一个矩阵，这样可以保证每一次迭代修复都能够减小

$$\min_{q^*, p^*} \sum_{(u,i) \in \mathcal{K}} (r_{ui} - q_i^T p_u)^2 + \lambda (\|q_i\|^2 + \|p_u\|^2)$$

直到收敛到一个标准。

ALS方法偏差处理与时间因素

- ▶ 用户打分倾向偏差Bias-bu
- ▶ 商品分数高低倾向偏差Bias-bi
- ▶ 动态因素：
 - 用户打分倾向会随时间改变-bu(t)
 - 商品分数高低倾向(流行度)会随时间改变-bi(t)
 - 用户兴趣特征会随时间改变-pu(t)
 - 商品特征可以认为近乎静态
- ▶ 引入动态因素的预测分：

$$\hat{r}_{ui}(t) = \mu + b_i(t) + b_u(t) + q_i^T p_u(t)$$

ALS implicit feedback处理

- ▶ 在隐反馈行为中，用户的喜好很难界定，只能陈述“可能喜欢这个商品”或者“可能不喜欢这个商品”
- ▶ 只能通过confidence信心值来估计用户的喜好度，信心值来源于用户对某个商品的重复行为
- ▶ Confidence: 用户对某个商品的单次行为可能源于很多因素而非喜欢，但用户对一个商品的多次重复行为更可能反映出用户的倾向。

$$\min_{p^*, q^*, b^*} \sum_{(u,i) \in K} c_{ui} (r_{ui} - \mu - b_u - b_i - p_u^T q_i)^2 + \lambda (\|p_u\|^2 + \|q_i\|^2 + b_u^2 + b_i^2)$$

c_{ui} : 用户对商品的信心度 *Confidence*

The end