

论文笔记——Slope One Predictors for Online Rating-Based Collaborative Filtering

Slope One 算法试图同时满足这样的5个目标

Slope One

Weighted Slope One

Bi-Polar Slope One

Slope One 算法试图同时满足这样的5个目标

1. 易于实现和维护
2. 运行时可更新的
3. 高效率的查询响应
4. 对初次访问者要求少
5. 合理的准确性

Slope One

其基本的想法来自于简单的一元线性模型 $w = f(v) = v + b$ 。

已知一组训练点 $(v_i, w_i)_{i=1}^n$ ，利用此线性模型最小化预测误差的平方和，我们可以获得

$b = \sum_i \frac{w_i - v_i}{n}$ ，b为 w_i 和 v_i 差值的平均值。

定义item i 相对于 item j 的平均偏差： $dev_{j,i} = \sum_{u \in S_{j,i}(\chi)} \frac{u_j - u_i}{card(S_{j,i}(\chi))}$

其中 $S_{j,i}(\chi)$ 表示同时对item i 和 j给予了评分的用户集合，而 $card(S)$ 表示集合S包含的元素数量。

可以用 $dev_{j,i} + u_i$ 获得用户u对 item j的预测值。当把所有这种可能的预测平均起来，可以得到：

$$P(u)_j = \frac{1}{card(R_j)} \sum_{i \in R_j} (dev_{j,i} + u_i)$$

其中 R_j 表示所有用户u 已经给予评分且满足条件 ($i \neq j$ 且 $S_{j,i}$ 非空) 的item集合。

对于足够稠密的数据集，我们可以使用近似 $\bar{u} = \sum_{i \in S(u)} \frac{u_i}{card(S(u))} \simeq \sum_{i \in R_j} \frac{u_i}{card(R_j)}$

把上面的预测公式简化为

$$P^{S1}(u)_j = \bar{u} + \frac{1}{card(R_j)} \sum_{i \in R_j} dev_{j,i}$$

Weighted Slope One

Slope One中在计算 item i 相对于 item j 的平均偏差 $dev_{j,i}$ 时没有考虑到使用不同的用户数量平均得到的 $dev_{j,i}$ ，其可信度不同。假设有2000个用户同时评分了 item j 和 k，而只有20个用户同时评分了 item j 和 l，那么显然获得的 $dev_{j,k}$ 比 $dev_{j,l}$ 更具有说服力。所以一个修正是对最终的平均使用加权：

$$P^{ws1}(u)_j = \frac{\sum_{i \in S(u) - \{j\}} (dev_{j,i} + u_i) c_{j,i}}{\sum_{i \in S(u) - \{j\}} c_{j,i}}, \text{ 其中 } c_{j,i} = card(S_{j,i}(\chi))$$

Bi-Polar Slope One

带权的slope-one算法很好地考虑了共同评分用户数的问题，但还有另外一个问题：好评和差评对用户的决策影响是不同的。很多的好评对用户的影响也不如少数的差评。因此这个算法先会计算用户对一个item的平均评分，然后将用户对item的评分集拆成两个，好评集和差评集：

$$S^{like}(u) = \{i \in S(u) | u_i > \bar{u}\}$$

$$S^{dislike}(u) = \{i \in S(u) | u_i < \bar{u}\}$$

类似地，可以定义对item i 和 j 具有相同喜好的用户集合：

$$S_{i,j}^{like} = \{u \in \chi | i, j \in S^{like}(u)\}$$

$$S_{i,j}^{dislike} = \{u \in \chi | i, j \in S^{dislike}(u)\}$$

利用上面的定义，我们可以使用下面的公式为（like或dislike的item）获得新的偏差值：

$$dev_{j,i}^{like} = \sum_{u \in S_{j,i}^{like}(\chi)} \frac{u_j - u_i}{card(S_{j,i}^{like}(\chi))}$$

这样可以计算从item i 计算得到的预测值：

$$p_{j,i}^{like} = dev_{j,i}^{like} + u_i \text{ 或者 } p_{j,i}^{dislike} = dev_{j,i}^{dislike} + u_i$$

最终 Bi-Polar Slope One 的预测公式为

$$P^{bps1}(u)_j = \frac{\sum_{i \in S^{like}(u) - \{j\}} p_{j,i}^{like} c_{j,i}^{like} + \sum_{i \in S^{dislike}(u) - \{j\}} p_{j,i}^{dislike} c_{j,i}^{dislike}}{\sum_{i \in S^{like}(u) - \{j\}} c_{j,i}^{like} + \sum_{i \in S^{dislike}(u) - \{j\}} c_{j,i}^{dislike}},$$

其中 $c_{j,i}^{like} = card(S_{j,i}^{like})$, $c_{j,i}^{dislike} = card(S_{j,i}^{dislike})$