

# 协同过滤推荐系统分析

王顺东, 雨爱华, 陈泽军

(西藏昌都军分区, 西藏昌都 854000)

**摘要:** 本文主要从基本思想、算法步骤等方面对基于用户的协同过滤推荐算法和基于项目的协同过滤推荐算法进行了详细介绍, 并对其存在的问题进行了总结。

**关键词:** 推荐系统; 协同过滤; 相似性

**中图分类号:** TP301.6 **文献标识码:** A **文章编号:** 1007-9599(2012)12-0152-01

随着互联网的飞速发展, 网络上的信息量剧增, 这使得人们很难从大量的信息中获得自己需要的信息, 这种现象降低了信息的使用率, 而这种信息过量但又不能被人们充分使用的现象叫做信息过载 (Information Overload)。而推荐系统 (Recommender System) 正是在这种背景下提出的, 成为解决信息过载问题的强大工具, 典型的电子商务系统如 epinions, eBay, Amazon 等都在一定程度上使用了推荐系统。在诸多的推荐方法中, 协同过滤推荐技术是最成功的技术之一。

协同过滤推荐技术的基本思想是根据用户的兴趣爱好信息, 查找与目标用户具有相似兴趣偏好的一些用户作为邻居, 然后根据这些邻居的兴趣信息为目标用户推荐其可能感兴趣的商品信息。其中, 基于用户的协同过滤推荐算法和基于项目的协同过滤推荐算法是目前应用最广泛的两种方法。

## 一、基于用户的协同过滤推荐算法

基于用户的协同过滤推荐算法是从用户间关系着手, 将根据兴趣相似用户感兴趣的项目推荐给目标用户。假设在推荐系统中, 存在  $m$  个用户和  $n$  个项目, 则用户-项目评分矩阵可以用  $m \times n$  阶矩阵  $R_{i,j}$  表示,  $R_{i,j}$  ( $1 \leq i \leq m, 1 \leq j \leq n$ ) 表示用户  $u_i$  对项目  $i_j$  的评分, 每个用户的评分信息均称为一条用户概貌。设定目标用户用  $u_a$ , 目标项目  $i_j$ 。采用基于用户的协同过滤推荐算法计算预测评分的过程如下:

(一) 寻找“邻居”用户: 首先根据公式 (1) 计算出目标用户  $u_a$  与其他所有用户的相似度; 然后根据相似度按照从大到小的顺序进行排序, 选取前  $k$  个用户作为目标用户  $u_a$  的邻居。

$$sim_{a,i} = \frac{\sum_{i_k \in I_{ai}} (R_{a,k} - \bar{R}_a)(R_{i,k} - \bar{R}_i)}{\sqrt{\sum_{i_k \in I_{ai}} (R_{a,k} - \bar{R}_a)^2} \sqrt{\sum_{i_k \in I_{ai}} (R_{i,k} - \bar{R}_i)^2}} \quad (1)$$

其中,  $sim_{a,i}$  表示目标用户  $u_a$  与邻居用户  $u_i$  的相似度;  $R_{a,k}$  和  $R_{i,k}$  分别表示  $u_a$  和  $u_i$  对项目  $i_k$  的评分;  $\bar{R}_a$  和  $\bar{R}_i$  分别表示  $u_a$  和  $u_i$  的平均评分;  $I_{ai}$  表示  $u_a$  和  $u_i$  的共同评分项目集。

(二) 进行预测: 根据  $k$  个邻居用户的评分信息, 根据公式 (2) 计算出目标用户  $u_a$  对目标项目  $i_j$  的预测评分。

$$P_{a,j} = \bar{R}_a + \frac{\sum_{u_k \in N(u_a)} (R_{k,j} - \bar{R}_k) \cdot sim_{a,k}}{\sum_{u_k \in N(u_a)} |sim_{a,k}|} \quad (2)$$

其中,  $P_{a,j}$  表示目标用户  $u_a$  对目标项目  $i_j$  的预测评分;  $N(u_a)$  表示目标用户  $u_a$  的邻居集;  $sim_{a,k}$  表示目标用户  $u_a$  与邻居用户  $u_k$  的相似度;  $R_{k,j}$  表示  $u_k$  对目标项目  $i_j$  的评分;  $\bar{R}_a$  和  $\bar{R}_k$  分别表示  $u_a$  和  $u_k$  的平均评分。

## 二、基于项目的协同过滤推荐算法

与基于用户的协同过滤推荐算法的基本思想不同, 基于项目的推荐算法是根据项目间的相似度选择出  $k$  个与目标项目最相似的项目, 然后根据目标用户对这些相似项目的评分信息计算出对目标项目的评分。其基本过程如下:

(一) 寻找“邻居”项目: 首先根据公式 (3) 计算出目标项目

$i_j$  与其他项目的相似度; 然后根据相似度按照从大到小的顺序进行排序, 选取前  $k$  个项目作为目标项目  $i_j$  的邻居。

$$sim_{i,j} = \frac{\sum_{u \in U_{ij}} (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U_{ij}} (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum_{u \in U_{ij}} (R_{u,j} - \bar{R}_j)^2}} \quad (3)$$

其中,  $sim_{i,j}$  表示项目  $i_i$  和目标项目  $i_j$  的相似度,  $U_{ij}$  表示对项目  $i_i$  和项目  $i_j$  共同评分的用户集合,  $R_{u,i}$  表示用户  $u_a$  对项目  $i_i$  的评分,  $\bar{R}_i$  和  $\bar{R}_j$  分别表示项目  $i_i$  和项目  $i_j$  的平均评分。

(二) 进行预测: 根据  $k$  个邻居项目的评分信息, 根据公式 (4) [3] 计算出目标用户  $u_a$  对目标项目  $i_j$  的预测评分。

$$P_{a,j} = \frac{\sum_{i_k \in S_j} R_{a,k} sim_{j,k}}{\sum_{i_k \in S_j} |sim_{j,k}|} \quad (4)$$

其中,  $P_{a,j}$  表示目标用户  $u_a$  对目标项目  $i_j$  的预测评分;  $S_j$  表示目标项目  $i_j$  的邻居集;  $sim_{j,k}$  表示目标项目  $i_j$  与邻居项目  $i_k$  的相似度;  $R_{a,k}$  表示  $u_a$  对项目  $i_k$  的评分。

## 三、存在的问题

虽然基于用户的协同过滤推荐算法和基于项目的协同过滤推荐算法作为推荐技术有其相当的应用, 但是随着电子商务规模的不断扩大, 用户和项目的数量迅速增长, 这两种算法也存在以下不足:

(一) 用户评分矩阵的极端稀疏性。随着电子商务系统规模的不断扩大, 用户数量和项目数量呈指数级增长, 将导致用户-项目评分矩阵极端稀疏。在这种数据量大而评分又稀疏的情况下, 一方面难以成功的定位邻居用户集, 另一方面进行相似性计算的代价很大, 直接导致推荐的准确度显著下降, 最后产生的推荐结果不理想。

(二) 冷启动问题。如果一个新项目没人评价它, 则这个项目得不到推荐, 推荐系统就失去了作用。同样, 如果一个新用户从未对系统中的项目进行评分或者评分信息非常少, 则系统无法获知他的兴趣, 也就无法对他进行推荐。

(三) 系统比较脆弱, 易被攻击。由于推荐系统自身所具有的开放性, 一些用户出于商业竞争等目的, 人为地向系统提供大量虚假的用户概貌, 企图使系统产生对他们有利的推荐结果, 降低用户对系统的目标和精度的信任, 从而达到其攻击的目的。

以上三个方面严重影响了系统推荐质量, 因此还需要进一步的研究。

## 参考文献:

- [1] Paul A C, Wolfgang N, Cristian Z. Preventing shilling attacks in online recommender systems. In Proc. 7th annual ACM international workshop on Web information and data management, Bremen, Germany, November 04, 2005, pp. 67 - 74.
- [2] Resnick P, Iacovou N, Sushak M, Bergstrom P, Riedl J. GroupLens: An open architecture for collaborative filtering of netnews. In Proc. of ACM 1994 Conf. on Computer Supported Cooperative Work. Chapel Hill, NC. Association of Computing Machinery. 1994, 175 - 186.
- [3] 邓爱林, 朱扬勇, 施伯乐. 基于项目评分预测的协同过滤推荐算法. 软件学报, 2003, 14(9): 1621 - 1628

# 协同过滤推荐系统分析

作者: [王顺东, 雨爱华, 陈泽军](#)  
作者单位: [西藏昌都军分区, 西藏昌都 854000](#)  
刊名: [计算机光盘软件与应用](#)  
英文刊名: [Computer CD Software and Applications](#)  
年, 卷(期): 2012(12)

## 参考文献(3条)

1. [Paul A C;Wolfgang N;Cristian Z Preventing shilling attacks in online recommender systems](#) 2005
2. [Resnick P;Iacovou N;Sushak M;Bergstrom P,Riedl J GroupLens: An open architecture for collaborative filtering of netnews](#) 1994
3. [邓爱林;朱扬勇;施伯乐 基于项目评分预测的协同过滤推荐算法](#) 2003(09)

本文链接: [http://d.wanfangdata.com.cn/Periodical\\_jsjgprjyyy201212104.aspx](http://d.wanfangdata.com.cn/Periodical_jsjgprjyyy201212104.aspx)