

# 改进的基于信任网络和随机游走策略的 评分预测模型

肖志宇<sup>1</sup>, 翟玉庆<sup>1, 2</sup>

(东南大学 1. 计算机科学与工程学院; 2. 计算机网络和信息集成教育部重点实验室, 江苏 南京 211189)

**摘要:** 为了提高推荐算法评分预测的准确度, 该文在 TrustWalker 模型的基础上, 提出了一个改进的基于信任网络和随机游走策略的评分预测模型——ReferentialUserWalker 模型。该模型通过随机游走策略, 利用信任网络中的信任朋友对目标物品或与目标物品相似的物品的评分进行评分预测, 并在信任网络中找到最可信的 TopN 评分参考用户, 同时引入信任度权重, 降低了噪声数据的影响。实验结果表明, 与 TrustWalker 模型相比, ReferentialUserWalker 模型的评分预测准确度有所提高。

**关键词:** 推荐; 信任网络; 随机游走; 评分预测; TrustWalker; 用户相似度; TopN; 评分参考用户

中图分类号: TP311 文章编号: 1005-9830(2015)05-0602-07

DOI: 10.14177/j.cnki.32-1397n.2015.39.05.015

## Improved rating prediction model basing on trust network and random walk strategy

Xiao Zhiyu<sup>1</sup>, Zhai Yuqing<sup>1, 2</sup>

(1. School of Computer Science and Engineering; 2. Key Lab of Computer Network & Information Integration, Southeast University, Nanjing 211189, China)

**Abstract:** In order to improve the accuracy of the rating prediction in recommender systems, the ReferentialUserWalker model based on TrustWalker is proposed here. The model is combined with the random walk strategy on the trust network and the item-based recommendation to improve the accuracy of rating prediction and to find the TopN trusted rating referential users associated with the trust weight to predict the rating. And then effect of noise data is reduced. The experiment results prove

收稿日期: 2015-03-17 修回日期: 2015-05-13

基金项目: 国家自然科学基金(6509000059)

作者简介: 肖志宇(1990-), 男, 硕士, 主要研究方向: 基于社交网络和信任模型的推荐系统, E-mail: xzy2502002@126.com; 通讯作者: 翟玉庆(1966-), 男, 教授, 主要研究方向: 人工智能及其应用, E-mail: yqzhai@seu.edu.cn。

引文格式: 肖志宇, 翟玉庆. 改进的基于信任网络和随机游走策略的评分预测模型[J]. 南京理工大学学报, 2015, 39(5): 602-608.

投稿网址: <http://zrxuebao.njust.edu.cn>

that the model in this paper has higher accuracy of rating prediction than TrustWalker.

**Key words:** recommender; trust network; random walk; rating prediction; TrustWalker; user similarity; TopN; rating referential users

对于日益增多的海量资源和信息,推荐技术可以满足用户从海量信息中筛选自己所需信息的需要。一种常用、实现过程简单、性能较好的推荐技术就是协同过滤<sup>[1]</sup>。评分预测,即利用用户对物品的历史评分信息来预测用户对没有接触过的物品的评分,是协同过滤推荐算法研究的核心,评分预测的准确性度量了其预测用户行为的能力<sup>[2]</sup>。数据稀疏性问题和冷启动问题是影响评分预测准确度的两大障碍因素。而基于信任的推荐系统<sup>[3]</sup>通过引入信任网络,很好地缓解了冷启动问题和数据稀疏性问题。

文献[4]提出的 TidalTrust 在信任网络中采用改进的广度优先搜索的策略,找出从源用户出发的最短路径长度内所有对目标物品评分过的用户后,迭代计算出源用户对目标物品的评分。最短路径长度设置为从源用户到第一个对目标物品评分过的用户之间的距离。由于 TidalTrust 只考虑最短路径长度内的其他用户对目标物品的评分,所以很多对评分预测有参考价值的用户的评分被忽略。

文献[5]提出的 TrustWalker 评分预测模型不仅考虑由信任网络中距离较近的朋友对与目标物品相似的物品的评分,路径长度也没有限制为 TidalTrust 中的最短路径长度。TrustWalker 模型主要由两部分组成:信任网络中的随机游走和相似物品的选择。通过在信任网络中随机游走选择相似物品的评分,从而避免在信任网络中游走太深。因此,TrustWalker 模型通过优先考虑信任网络中用户对目标物品的评分从而保证了评分的准确性,又通过考虑与目标物品相似的物品的评分提高了推荐的覆盖率。

本文中评分预测的目标用户用符号  $u_0$  表示,目标物品用  $i_0$  表示,目标用户  $u_0$  对目标物品  $i_0$  的预测评分用  $\hat{r}_{u_0 i_0}$  表示。另外,文中的信任网络中每个节点代表相应的用户,每条边代表用户之间的信任关系,每条边的信任值都是 1(即,用户之间要么信任(有边相连),要么没有信任(无边相连))。

## 1 TrustWalker 模型

TrustWalker 模型主要思想是将基于信任网络的推荐算法和基于项目的协同过滤算法进行了结合,不仅考虑信任网络中的用户对于目标物品  $i_0$  的评分,也考虑他们对于那些和  $i_0$  相似的物品的评分,而且随着随机游走深度的增加,当前停留节点离用户  $u_0$  越远,以相似物品评分代替  $i_0$  的评分的概率也增加。

### 1.1 一趟随机游走过程

从目标用户  $u_0$  出发,对目标物品  $i_0$  进行预测,在信任网络中执行随机游走,每一步游走都会抵达一个节点用户。假如在第  $k$  步随机游走(即游走深度为  $k$ )时抵达用户  $u$ ,若  $u$  对  $i_0$  有评分,则停止该趟随机游走,并且返回评分  $r_{u i_0}$ ,否则执行如下步骤:(1)以  $\phi_{u i_0 k}$  的概率在用户  $u$  处停止该趟游走,并且在  $u$  的评分物品集中随机选择一个与  $i_0$  相似的物品  $j$ ,返回评分  $r_{u j}$ ;(2)以  $1 - \phi_{u i_0 k}$  的概率继续在信任网络中随机游走,第  $k+1$  步跳转到用户  $u$  的任意一个直接信任节点用户上。

#### 1.1.1 选择下一步跳转节点的概率

假设  $TU_u$  是用户  $u$  的直接信任用户集合, $S_u$  是在下一步的随机游走中从用户  $u$  的直接信任邻居结合中选择某一个用户  $v$  的随机变量,则在节点  $u$  时选择用户  $v$  作为下一步跳转节点的概率计算公式为

$$P(S_u = v) = \frac{1}{|TU_u|} \quad (1)$$

从式(1)可以看出,TrustWalker 模型是以等概率来选择下一步跳转节点的。

#### 1.1.2 选择相似物品的方法

当游走到第  $k$  步决定以  $\phi_{u i_0 k}$  的概率在用户  $u$  处停止该趟游走时,要在  $u$  的评分物品中随机选择一个和  $i_0$  相似的物品  $j$ 。假设  $Y_u$  是从用户  $u$  的评分记录中选择一个物品  $j$  的随机变量,则选择物品  $j$  作为相似物品的概率为

$$P(Y_u=j) = \frac{\text{sim}(i_0, j)}{\sum_{l \in RI_u} \text{sim}(i_0, l)} \quad (2)$$

式中:  $\text{sim}(i_0, j)$  表示物品  $i_0$  与  $j$  的物品相似度。

### 1.1.3 物品相似度计算

TrustWalker 物品相似度的计算类似于文献 [6] 中的方法, 首先计算两个物品之间相似度的改进余弦

$$\text{corr}(i, j) = \frac{\sum_{u \in UC_{ij}} (r_{ui} - \bar{r}_u)(r_{uj} - \bar{r}_u)}{\sqrt{\sum_{u \in UC_{ij}} (r_{ui} - \bar{r}_u)^2} \sqrt{\sum_{u \in UC_{ij}} (r_{uj} - \bar{r}_u)^2}} \quad (3)$$

式中:  $UC_{ij}$  是指既对物品  $i$  和  $j$  均有评分的用户集合。考虑  $UC_{ij}$  的大小, 从而物品相似度的最终计算公式为

$$\text{sim}(i, j) = \frac{1}{1 + e^{-|UC_{ij}|/2}} \times \text{corr}(i, j) \quad (4)$$

在选择相似物品时, 由于  $\text{corr}(i, j) \in [-1, 1]$ , 物品相似度因此也在  $[-1, 1]$  之间, TrustWalker 模型只考虑相似度大于 0 的物品作为选择。

### 1.1.4 一趟随机游走的终止条件

只要符合如下 3 个条件之一, 单趟随机游走即终止:

(1) 游走过程中抵达的某个节点用户  $u$  对目标项目  $i_0$  有评分, 返回  $r_{ui_0}$ , 停止这一趟随机游走;

(2) 在节点  $u$  处在停止概率  $\phi_{ui_0k}$  下, 从用户  $u$  的评分记录中选了一个与  $i_0$  相似的物品  $j$ , 返回  $r_{uj}$ , 该趟随机游走终止;

(3) 超过预先设置的随机游走的最大深度。

其中停止概率  $\phi_{ui_0k}$  既与  $u$  评分过的物品与  $i_0$  的最大相似度有关, 也与当前游走的深度  $k$  有关。物品相似度越高, 深度越大, 停止的概率也应该越大。因此, 停止概率的计算方法为

$$\phi_{ui_0k} = \max_{j \in RI_u} \text{sim}(i_0, j) \times \frac{1}{1 + e^{-k/2}} \quad (5)$$

式中:  $\text{sim}(i_0, j)$  为式 (4) 中物品相似度计算方法,  $\max_{j \in RI_u} \text{sim}(i_0, j)$  为用户  $u$  评分过的物品与目标物品  $i_0$  相似度中的最大值,  $k$  代表当前随机游走的深度。

### 1.2 全局随机游走终止条件

首先计算所有单趟随机游走返回分数的方差, 假设第  $m$  趟随机游走返回的评分为  $r_m$ , 前  $n$  趟游走过程返回的评分平均值为  $\bar{r}$ , 则方差  $\sigma_n^2$  计

算方法为

$$\sigma_n^2 = \frac{\sum_{m=1}^n (r_m - \bar{r})^2}{n} \quad (6)$$

若  $\sigma_n^2$  收敛即  $|\sigma_{n+1}^2 - \sigma_n^2| \leq \varepsilon$ , 全局随机游走过程终止。 $\varepsilon$  为预先设置的一个很小的阈值, 如 0.000 1。

### 1.3 评分预测

TrustWalker 模型最终预测目标用户  $u_0$  对目标  $i_0$  的评分是各趟随机游走过程中所返回评分信息的融合, 计算方法为

$$\hat{r}_{u_0 i_0} = \sum P(XY_{u_0 i_0} = (u, j)) r_{uj} \quad (7)$$

式中:  $XY_{u_0 i_0}$  是在计算目标用户  $u_0$  对于物品  $i_0$  预测评分过程中抵达节点  $u$  且从  $u$  的评分记录中选择物品  $j$  的评分的随机变量。

## 2 ReferentialUserWalker 模型

本文提出的 ReferentialUserWalker 模型借鉴 TrustWalker 模型框架, 并有如下改进:

(1) TrustWalker 模型是以等概率从当前节点的直接信任用户集合中挑选下一个跳转节点, 但我们认为用户之间的相似度也应该作为选择下一步跳转节点时需要考虑的因素, 也就是说与当前节点具有更高的相似度的节点, 在游走的下一步中应该以更大的概率被选择;

(2) 每趟随机游走停留在当前节点的概率不仅考虑当前节点评分过的物品与目标物品的物品相似度的最大值和游走深度, 也应该考虑当前节点和目标用户  $u_0$  的用户相似度;

(3) 一些在随机游走过程中停留概率较小的节点的评分信息有可能是噪声数据, 所以本文重新设置了全局游走的停止条件, 每趟随机游走只返回一个用户, 并且计算评分时, 只选取最可信的 TopN 用户对目标物品  $i_0$  的评分或与  $i_0$  相似的物品评分, 同时加入信任度权重进一步提高评分预测的准确度。

通过将用户相似度作为选择下一跳节点的重要因素, 能有更高的概率跳转到对预测评分有重要作用的用户; 同样在每个节点停止的条件不仅考虑目标物品  $i_0$  的评分, 也考虑与  $i_0$  相似物品的评分, 游走过程中有更多的评分信息, 从而也保证了模型推荐的覆盖率; 在评分时只考虑 TopN 用户的评分, 并考虑相应用户的信任度权重, 从而使

评分更为准确。

## 2.1 一趟随机游走过程

从目标用户  $u_0$  出发,对目标物品  $i_0$  进行预测,在信任网络中执行随机游走,每一步游走都会抵达一个节点用户。假如在第  $k$  步随机游走时抵达用户  $u$ ,这时有两种选择:(1)在一定情况下,随机游走停止在用户  $u$ ,并根据具体情况返回相应的值;(2)随机游走继续,并从用户  $u$  的直接信任邻居中按照概率  $P(S_u=v)$  选择一个用户  $v$  作为随机游走下一步的节点。

### 2.1.1 选择下一步跳转节点的概率

假设  $TU_u$  是用户  $u$  的直接信任用户集合,  $S_u$  是在下一步的随机游走中从用户  $u$  的直接信任邻居结合中选择某一个用户  $v$  的随机变量,则在节点  $u$  时选择用户  $v$  作为下一步跳转节点的概率计算为

$$P(S_u=v) = \frac{\text{usersim}^+(u, v)}{\sum_{\omega \in TU_u} \text{usersim}^+(u, \omega)} \quad (8)$$

从节点  $u$  选择继续游走必须从用户  $u$  直接信任邻居中选取一个作为下一步的节点。 $\text{usersim}^+(u, v)$  代表用户  $u$  与用户  $v$  的用户相似度,下文具体介绍用户相似度计算方法。从式(8)可以看出,选择用户  $u$  的某个直接信任邻居  $v$  作为游走的下一步的节点概率,是用户  $u$  和用户  $v$  之间相似度占用户  $u$  与所有直接信任邻居用户相似度之和的比值。

### 2.1.2 用户相似度计算

本文计算用户相似性是利用用户对共同评分过的物品的评分,采用文献[7]中改进的余弦相似度来计算相似性

$$u\_sim(u, v) = \frac{\sum_{i \in IC_{u,v}} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in IC_{u,v}} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in IC_{u,v}} (r_{v,i} - \bar{r}_v)^2}} \quad (9)$$

被用户  $u$  和  $v$  都评分过的共同物品集合大小也十分重要,共同物品集合越大,用户相似度应该越高,所以用户相似度计算公式为

$$\text{usersim}(u, v) = \frac{1}{1 + e^{-\frac{|IC_{u,v}|}{2}}} \times u\_sim(u, v) \quad (10)$$

虽然有评分的物品总数很多,但绝大多数用户之间的共同评分的物品数很少,这就导致许多用户之间的  $\text{usersim}(u, v)$  为负数或零,所以如果

仅按照式(10)计算下一步的随机游走跳转到每个节点的概率,大部分节点被选择到的概率都为零,这就阻碍了随机游走的进行,为使式(10)的计算结果满足概率分布,将其映射到  $[0, 1]$  区间

$$\text{usersim}^+(u, v) = \frac{1 + \text{usersim}(u, v)}{2} \quad (11)$$

### 2.1.3 物品相似度计算

本文物品相似度计算方法与 TrustWalker 中相同,计算两个物品之间相似度的改进余弦计算公式为

$$i\_sim(i, j) = \frac{\sum_{u \in UC_{i,j}} (r_{u,i} - \bar{r}_u)(r_{u,j} - \bar{r}_u)}{\sqrt{\sum_{u \in UC_{i,j}} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{u \in UC_{i,j}} (r_{u,j} - \bar{r}_u)^2}} \quad (12)$$

同样考虑对物品  $i, j$  都有评分的共同用户集合大小后,物品相似度最终计算公式为

$$\text{itemsim}(i, j) = \frac{1}{1 + e^{-\frac{|UC_{i,j}|}{2}}} \times i\_sim(i, j) \quad (13)$$

### 2.1.4 一趟随机游走的终止条件

一趟随机游走的第  $k$  步游走到用户  $u$  时,会有以下 4 种终止情况:

(1) 若节点  $u$  对目标物品  $i_0$  有评分,则返回用户  $u$ ,并终止这趟随机游走;

(2) 当游走步数大于最大步数时,终止这趟随机游走,本趟随机游走不成功,结果返回空(为了防止一直选择继续随机游走,预先设置最大游走步数,根据“六度分隔理论”<sup>[8,9]</sup>,设置最大游走步数为 6);

(3) 节点  $u$  对目标物品  $i_0$  没有评分,有  $\phi_{u_0, u, i_0, k}$  的概率停留在节点  $u$ ,根据“轮盘赌”随机选择模型,若随机选择停止,则返回用户  $u$ ,并终止这趟随机游走;否则,继续向下游走;

(4) 在节点  $u$  若选择继续游走,但是  $u$  没有直接信任邻居,则终止这趟随机游走,本趟随机游走不成功,返回空。

停止概率  $\phi_{u_0, u, i_0, k}$  应该与用户  $u$  评分过的物品与  $i_0$  的相似度有关,考虑用户  $u$  评分过的物品中与  $i_0$  物品相似度的最大值作为停止概率中的重要因素。另外,当前用户  $u$  与目标用户  $u_0$  的用户相似度也是一个重要因素,用户相似度越高,该用户的评分越有参考价值,停止概率应该越大。

在信任网络中离  $u_0$  越近的用户评分越可靠,所以游走步数也是停止概率考虑的重要因素。

在信任网络中游走的步数越大,随机游走停止的概率应该越大。

综合以上因素  $\phi_{u_0 \mu i_0 k}$  的计算公式为

$$\phi_{u_0 \mu i_0 k} = \max_{j \in RI_u} \{ \text{itemsim}(i_0 j) \} \times \text{usersim}^+(u, \mu_0) \times \frac{1}{1 + e^{-\frac{k}{2}}} \quad (14)$$

式中:  $\text{itemsim}(i, j)$  为式(13)中物品相似度的计算方法,  $\max_{j \in RI_u} \{ \text{itemsim}(i, j) \}$  代表用户  $u$  评分过的物品与目标物品  $i_0$  相似度中的最大值,  $\text{usersim}^+(u, \mu_0)$  为式(11)中的计算方法,  $k$  代表当前随机游走的深度。

## 2.2 全局游走终止条件

ReferentialUserWalker 模型在执行完成多趟随机游走过程后,会得到一个评分参考用户列表。为了确定 TopN 评分参考用户列表是否已经足够稳定并且符合评分预测的需要,设计了基于每  $I$  趟随机游走的 ReferentialUserWalker 模型全局游走的终止条件。

随机游走过程中,在每进行  $I$  趟随机游走后,ReferentialUserWalker 模型会通过计算比较当前 TopN 评分参考用户列表中的用户权值向量和当前  $I$  趟随机游走之前的用户权值向量。TopN 评分参考用户列表中每个用户的权值,用符号  $\omega_u$  表示,计算方法是在当前 TopN 评分参考用户列表中通过返回用户  $u$  的频次除以所有用户的频次之和。

在进行每  $I$  趟随机游走之前的 TopN 评分参考用户列表用  $L'$  表示,其中每个用户  $u \in L'$  都有一个权值  $\omega'_u$ 。每  $I$  趟随机游走之后,该列表会根据出现的频次更新为一个新列表  $L$ ,其中每个用户  $u \in L$  都有一个权值  $\omega_u$ 。比较两个列表的权值向量的方法为

$$\text{diff}(L, L') = \sqrt{\sum_{u \in L \cap L'} (\omega_u - \omega'_u)^2 + \sum_{u \in L, u \notin L'} 1} \quad (15)$$

式(15)对于两个列表中共同出现的用户,计算它们之间权值向量的“欧几里德距离”。也可能出现两个列表中拥有不同用户的情况,因为 TopN 参考用户列表中每个用户的权值都在  $[0, 1]$  之间,所以  $(\omega_u - \omega'_u)^2$  项的值也在  $[0, 1]$  之间。因此,对于每个在新列表  $L$  中出现但在  $L'$  中没出现的用户,ReferentialUserWalker 模型会分配最大距离值 1 作为惩罚因子。

综上,ReferentialUserWalker 模型全局随机游

走的终止条件为  $\text{diff}(L, L') \leq \varepsilon$ 。如果不成功的随机游走趟数(即返回用户为空的随机游走趟数)达到设置的最大阈值(本文设置为 10 000),全局随机游走立刻终止,使用当前 TopN 用户列表进行评分预测。当全局随机游走终止后,如用户列表为空,则认为 ReferentialUserWalker 模型无法对这个用户物品对进行评分预测。

## 2.3 评分预测

每趟随机游走返回的结果会放入一个用户列表,每个用户对应一个出现的频次即随机游走返回该用户的次数,根据这样一个频次将用户从高到低进行排序,选出返回次数最多的 TopN 用户作为评分参考用户列表。以评分参考用户列表中每个用户出现的次数占列表中所有用户出现总次数的比例作为最终评分计算时每个用户的权重。若列表中用户对目标物品  $i_0$  有评分,则使用该评分作为参考评分;若对目标物品  $i_0$  没有评分,则参考评分为该用户评分记录中与  $i_0$  物品相似度大于 0 的所有物品的评分加权求和,每个物品对应的权重为该物品相似度占所有物品相似度之和的比值(相似度小于 0 的物品不考虑)。综上,假设  $RI_u^+$  为用户  $u$  评分记录中与  $i_0$  物品相似度大于 0 的物品集合,那么用户  $u$  对  $i_0$  的参考评分  $r_{u \mu_0 i_0}$  计算公式为

$$r_{u \mu_0 i_0} = \frac{\sum_{i \in RI_u^+} \text{itemsim}(i, i_0) \times r_{u i}}{\sum_{i \in RI_u^+} \text{itemsim}(i, i_0)} \quad (16)$$

将信任该用户的用户数作为该用户参考评分权重所考虑的因素,称其为信任度权重。在对用户  $u$  的一个未评分项目  $i$  进行评分预测时会引入该信任度权重。认为被更多其他用户所信任的用户的参考评分越可靠,信任度权重也越高。所以信任度权重的计算方法为

$$\text{Confidence}_u = \frac{1}{1 + e^{\frac{-|Tn_u|}{2}}} \quad (17)$$

式中:  $|Tn_u|$  代表信任用户  $u$  的用户数。在求得信任权重后,最终评分预测经 TopN 评分参考用户列表中每个用户的权重和参考评分加权求和得出,用户  $u_0$  对目标物品  $i_0$  的预测评分计算公式为

$$\hat{r}_{u_0 i_0} = \sum_{u \in L} (\omega_u \times r_{u \mu_0 i_0} \times \text{Confidence}_u) \quad (18)$$

式中:  $L$  为 TopN 评分参考用户列表,  $\omega_u$  为用户  $u$  的权重,  $r_{u \mu_0 i_0}$  为用户  $u$  对目标物品  $i_0$  的参考评分,  $\text{Confidence}_u$  为用户  $u$  的信任度权重。

### 3 实验分析

#### 3.1 实验设计

本文实验使用的 Epinions 数据集<sup>[10]</sup>既包含用户之间的信任网络,也包含每个用户对物品的评分信息。数据集中的评分信息为 [1, 5] 区间的离散值。数据集有 49 K 个用户,评分过的物品数有 139 K 个,用户之间的信任关系有 487 K 个,评分数量为 664 K。

本文使用评分预测中通常使用的留一法 (Leave-one-out) 方法来评估模型,保留一条评分信息,利用剩余的评分信息和信任网络来预测该评分。同时,若抽取出一条测试数据,该测试数据中的用户既没有信任网络,又没有评分信息,则去除这样的测试数据,因为此情形无法预测评分。本文使用的 Epinions 数据集十分稀疏,有评分记录的用户数量为 40 K,其中评分少于 5 条的用户却有将近 17 K,文献 [11] 将这样的用户看作冷启动用户。存在大量冷启动用户会对评分预测的准确度影响较大。为此,分别在全集用户集和冷启动用户集两个测试集上单独进行实验,以比较 ReferentialUserWalker 模型与 TrustWalker 模型。

#### 3.2 实验参数

在 ReferentialUserWalker 模型中设置终止条件计算的趟数间隔  $I$  为 1 000,每趟游走最大深度为 6,与文献 [12] 中协同过滤中设置最相似的  $k$  个物品相同,本文 TopN 值设为 50,终止条件  $\varepsilon = 0.005$ ,最大游走失败次数为 10 000。

#### 3.3 实验评测指标

评分预测的预测准确度一般通过均方根误差 (RMSE) 来计算,如文献 [4, 6, 10] 中的研究方法,均方根误差在测试集  $T$  中的计算方法为

$$RMSE = \sqrt{\frac{\sum_{(u,i) \in T} (r_{u,i} - \hat{r}_{u,i})^2}{|T|}} \quad (19)$$

对于测试集中每个用户物品对,不一定都能预测出评分,所以用另一个指标覆盖率 (Coverage) 来表示评分预测模型能够预测出评分的用户物品对占测试集的比例。覆盖率也能体现出评分预测模型一方面的能力。ReferentialUserWalker 模型中如果全局随机游走停止后没有评分参考用户返回,则认为评分预测模型无法覆盖到该用户物品对。使用  $Success\_T$  表示测试集中能够预测评分的用

户物品对集合,则覆盖率 (Coverage) 的计算方法

$$Coverage = \frac{|Success\_T|}{T} \quad (20)$$

还可以通过计算 FMeasure 将均方根误差 (RMSE) 和覆盖率 (Coverage) 转化为一个值。为此必须将 RMSE 转化为一个在 [0, 1] 区间的值。由于数据集中的评分都在 [1, 5] 区间,所以 RMSE 最大可能的值为 4。为此,定义准确度 (Precision)

$$Precision = 1 - \frac{RMSE}{4} \quad (21)$$

FMeasure 的计算方法为

$$FMeasure = \frac{2 \times Precision \times Coverage}{Precision + Coverage} \quad (22)$$

#### 3.4 实验结果

分别在全集用户集和冷启动用户集上对两种模型的准确率、用户覆盖率和 FMeasure 指标进行比较。图 1 显示了在冷启动用户集上 3 种指标的比较结果。由图 1 发现,ReferentialUserWalker 模型与 TrustWalker 模型推荐的覆盖率基本持平,而其准确率则比 TrustWalker 模型提高 0.2%,从而其 FMeasure 指标也相应提高 0.1%。

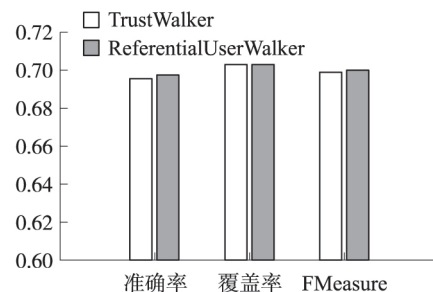


图 1 冷启动用户集实验结果

图 2 显示了在全集用户集上 3 种指标的比较结果。由图 2 可见,在全集用户测试集上,ReferentialUserWalker 模型比 TrustWalker 的准确率提升 0.5%,覆盖率提升 1%,FMeasure 指标提高近 1%。

通过以上实验结果分析可知,ReferentialUserWalker 模型通过只考虑 TopN 评分用户的评分提高了准确率,通过改进全局随机游走停止条件和用户的信任度权重提高了覆盖率。而对比冷启动用户集和全集用户集的数据可以发现,两种模型在全集用户测试集各个指标上都要优于冷启动用户测试集的实验结果,这是因为全集用户集比冷启动用户集包含更多的评分信息。

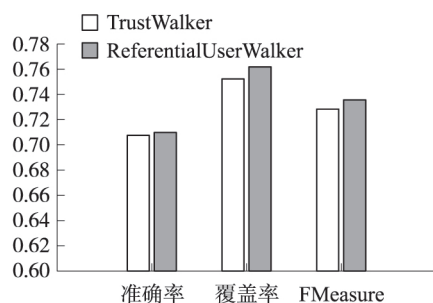


图 2 全集用户集实验结果

## 4 结论

本文提出的 TrustWalker 的改进模型——ReferentialUserWalker 模型,通过将用户相似度作为选择下一跳节点的重要因素,获得了更高的概率跳转到具有评分参考价值的用户;同样在每个节点停止的条件不仅考虑目标物品评分,也考虑相似物品的评分,从而保证了模型对推荐的覆盖率;在评分时只考虑 TopN 评分用户和用户的信任度权重,使评分更为准确。本文实验表明,相对于 TrustWalker 模型,ReferentialUserWalker 模型的评分预测准确度有所提高。

但是文献[13, 14]中认为在推荐系统的实际应用中,给用户推荐一个物品列表比预测用户对一个物品的评分更具有实际意义,更符合用户需求,因此需要将本文的模型扩展成 TopN 物品的推荐模型。另外,本文中的信任网络是一个二值信任网络,用户之间信任关系只有 0 或者 1,然而每个用户对其他用户的信任程度应该因人而异,所以还应考虑信任关系的取值及其动态变化。

### 参考文献:

- [1] Zan Huang, Zeng Daniel, Chen Hsinchun. A comparison of collaborative-filtering recommendation algorithms for e-commerce [J]. IEEE Intelligent Systems, 2007, 22(5): 68-78.
- [2] 项亮. 推荐系统实践 [M]. 北京: 人民邮电出版社, 2012.
- [3] Massa P, Bhattacharjee B. Using trust in recommender systems: An experimental analysis [M]. Berlin, Germany: Springer, 2004.
- [4] Golbeck J. Computing and applying trust in Web-based social networks [D]. Washton D C, US: University of Maryland College Park, 2005.
- [5] Jamali M, Ester M. TrustWalker: A random walk model for combining trust-based and item-based recommendation [A]. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [C]. Pairs, France: ACM, 2009: 397-406.
- [6] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms [A]. Proceedings of the 10th International Conference on World Wide Web [C]. Hong Kong, China: ACM, 2001: 285-295.
- [7] Jannach D, Zanker M, Felfernig A, et al. Recommender systems: An introduction [M]. New York, US: Cambridge University Press, 2010.
- [8] Watts D J. Six degrees: The science of a connected age [M]. New York, US: WW Norton & Company, 2004.
- [9] Jamali S M. Probabilistic models for recommendation in social networks [D]. Burnaby, Canada: School of Computing Science, Simon Fraser University, 2013.
- [10] Massa P. Epinions datasets [EB/OL]. [http://www.trustlet.org/wiki/Epinions\\_dataset](http://www.trustlet.org/wiki/Epinions_dataset), 2011-11-04.
- [11] Massa P, Avesani P. Trust-aware recommender systems [A]. Proceedings of the 2007 ACM Conference on Recommender Systems [C]. Minneapolis, US: ACM, 2007: 17-24.
- [13] 张富国. 基于社交网络的个性化推荐技术 [J]. 小型微型计算机系统, 2014, 35(7): 1470-1476.  
Zhang Fuguo. Survey of online social network based personalized recommendation [J]. Journal of Chinese Computer Systems, 2014, 35(7): 1470-1476.
- [14] 谭学清, 黄翠翠, 罗琳. 社会化网络中信任推荐研究综述 [J]. 现代图书情报技术, 2014, 30(11): 10-16.  
Tan Xueqing, Huang Cuicui, Luo Lin. A review of research on trust recommendation in social networks [J]. New Technology of Library and Information Service, 2014, 30(11): 10-16.