

# 基于二分图网络的总体多样性增强推荐算法\*

张骏<sup>1,2</sup>, 丁艳辉<sup>1,2</sup>, 金连旭<sup>1,2</sup>, 赵文朋<sup>1,2</sup>

(1. 山东师范大学信息科学与工程学院, 济南 250014; 2. 山东省物流优化与预测工程技术研究中心, 济南 250014)

**摘要:** 针对传统推荐算法过于强调推荐准确率而造成推荐系统长尾现象加剧问题, 提出一种基于二分图网络的总体多样性增强推荐算法。首先, 利用现有推荐算法生成的预测评分构建用户候选推荐列表, 进而构建二分图网络模型; 其次, 设定项目容量对热门项目的推荐次数予以限制; 最后, 结合推荐增广路生成最终推荐列表。与现有的推荐多样性增强算法在真实电影评分数据集上进行实验对比, 实验结果表明, 该算法在保证推荐准确率的同时能有效提高推荐的总体多样性。

**关键词:** 推荐系统; 总体多样性; 二分图

中图分类号: TP391

文献标志码: A

文章编号: 1001-3695(2018)06-1628-03

doi:10.3969/j.issn.1001-3695.2018.06.006

## Enhanced algorithm for recommendation aggregate diversity based on bipartite graph networks

Zhang Jun<sup>1,2</sup>, Ding Yanhui<sup>1,2</sup>, Jin Lianxu<sup>1,2</sup>, Zhao Wenpeng<sup>1,2</sup>

(1. School of Information Science & Engineering, Shandong Normal University, Jinan 250014, China; 2. Shandong Provincial Logistics Optimization & Predictive Engineering Technology Research Center, Jinan 250014, China)

**Abstract:** The traditional recommendation algorithm emphasizes too much on the accuracy of the recommendation, resulting in the “long-tail” phenomenon of the recommendation system intensified. Therefore, this paper proposed an enhanced algorithm for recommendation aggregate diversity based on the bipartite graph networks. Firstly, it constructed candidate user recommendation list based on predictive scores generated by the existing recommendation algorithm, and then constructed the bipartite graph network model. Secondly, set the item-capacity to limit the number of recommendations of popular items. Finally, it generated the final recommendation result in conjunction with the recommendation augmenting path. It compared the experimental results with that of the existing recommendation aggregate diversity algorithm on real-world movie rating datasets. The experimental results show that the proposed algorithm can effectively guarantee the accuracy of the recommendation results as well as improve the aggregate diversity of the recommendation.

**Key words:** recommendation system; aggregate diversity; bipartite graph

个性化推荐技术提供智能化、个性化机制帮助用户从海量数据中获取到其感兴趣的信息, 已成为最有效的信息过滤技术之一<sup>[1,2]</sup>。经过十几年的发展, 已广泛应用于电子商务、影视推荐、社交推荐等领域<sup>[3,4]</sup>。近年来, 越来越多的学者意识到推荐的准确率不应成为评价推荐质量的唯一标准。推荐的多样性、新颖性、惊喜度、覆盖率等因素对于推荐系统的评价同样具有重要意义<sup>[5]</sup>。其中, 推荐多样性的重要性已在若干研究中被强调<sup>[6,7]</sup>。通常推荐的多样性包括个体多样性与总体多样性。个体多样性从单个用户的角度考虑, 其目的在于为用户推荐一系列符合用户兴趣且彼此相似度低的物品, 从而开阔用户视野。总体多样性从系统角度考虑, 提高总体多样性的目的在于提高系统中冷门物品的推荐能力, 减弱系统长尾现象。本文针对推荐系统总体多样性进行研究, 提出一种基于二分图网络的总体多样性增强推荐算法 (AD-Improved 算法)。

优秀的推荐系统需要同时兼顾准确率与多样性两个推荐指标。然而推荐准确率与多样性彼此相互制约。部分学者采用启发式策略, 以牺牲一定程度准确率为代价提高推荐多样

性。Adomavicius 等人<sup>[8]</sup>将推荐系统建模为最大二分图匹配问题进行求解, 最大程度地提高了项目的覆盖率, 然而针对长尾现象并未有效解决。Adomavicius 等人<sup>[9]</sup>提出一种重排名算法, 通过对项目进行重排序以提高冷门项目的排名, 进而提高总体多样性, 具有较低的算法复杂度, 然而缺少全局意识。刘慧婷等人<sup>[10]</sup>从全局角度考虑, 提出一种基于推荐期望的多样性优化算法, 通过控制全体物品的推荐期望来提高推荐总体多样性。本文算法在现有推荐算法生成项目预测评分的基础上, 针对每位用户筛选预测评分大于阈值的项目构建候选推荐列表, 进而在候选推荐列表的基础上构建二分图网络模型, 通过对项目的推荐次数加以限制并结合推荐增广路信息, 生成最终的推荐列表。本文算法具有以下优势: a) 该算法可对多数现有推荐算法进行总体多样性二次优化; b) 通过对评分阈值进行设置, 可以灵活地在多样性与准确率两者间取折中。

### 1 基于二分图网络的总体多样性增强推荐算法

二分图是图论中一种特殊模型, 其在复杂网络理论研究与

收稿日期: 2017-01-15; 修回日期: 2017-03-14 基金项目: 国家自然科学基金青年基金资助项目 (61303007)

作者简介: 张骏 (1991-), 男, 山东烟台人, 硕士研究生, 主要研究方向为推荐系统、数据挖掘 (zhangjun\_pl@163.com); 丁艳辉 (1981-), 男, 山东济南人, 副教授, 博士, 主要研究方向为 Web 数据集成、医学图像分析; 金连旭 (1990-), 男, 山东济南人, 硕士研究生, 主要研究方向为推荐系统、数据挖掘; 赵文朋 (1991-), 男, 山东济宁人, 硕士研究生, 主要研究方向为数据挖掘。

实际应用中都具有非常重要的意义。二分图最大匹配问题作为二分图的一个重要研究问题,已具有成熟的求解算法。匈牙利算法通过寻找增广路并置换增广路中匹配边与非匹配边实现二分图的最大匹配,该算法相比最大流算法具有更高的效率。推荐系统的二分图区别于传统二分图,每个项目都可被推荐给多个用户。因此,推荐二分图中无法直接使用匈牙利算法。在用户数量与项目推荐数量固定的推荐系统中,减少热门项目的推荐次数,将会提高冷门项目的推荐机会,进而提高系统总体多样性。本文对项目被推荐的次数予以限制,并借鉴匈牙利算法寻找增广路进而置换增广路中匹配边与非匹配边思路,增加冷门项目的推荐次数。其次设置了评分阈值,二分图中用户对非匹配边连接项目的预测评分仍保持在较高水平。推荐系统将非匹配边连接的项目推荐给用户,对推荐准确率影响较小。

### 1.1 相关定义

假设由  $n$  名用户 user 构成的用户集合为  $U = \{u_1, u_2, \dots, u_n\}$ ,  $m$  个项目 item 构成的项目集合为  $I = \{i_1, i_2, \dots, i_m\}$ 。用户对项目的评分可由  $n \times m$  维矩阵  $R$  表示,其中矩阵的每一项  $R(u, i)$  表示用户  $u$  对项目  $i$  的真实评分。用户对未评分项目的预测评分由  $n \times m$  维矩阵  $R^*$  表示,其中矩阵的每一项  $R^*(u, i)$  表示用户  $u$  对项目  $i$  的预测评分。预测评分值可由 User-based CF、Item-based CF、Matrix Factorization CF 等现有算法产生。定义用户推荐列表为  $Tu$ , 推荐列表长度为  $N$ , 定义项目可被推荐的次数为项目容量,用  $IC$  表示。推荐初始时,所有项目的项目容量相同。

**定义 1** 候选推荐列表。设置评分阈值  $Tr$ , 用户  $u$  的候选推荐列表:  $Cu = \{item | item \in I, u \in U \text{ and } R^*(u, item) > Tr\}$ , 用户候选列表中的项目可视为用户喜爱的项目。

**定义 2** 推荐二分图。  $G = \langle U, I, E \rangle$ ,  $U$  为用户节点的集合,  $I$  为项目节点的集合,  $E$  为用户  $u$  与项目  $i$  间关系的集合。其中, 连接用户  $u$  与其  $Cu$  中项目的边称为匹配边, 连接用户  $u$  与其  $Tu$  中项目的边称为非匹配边。

**定义 3** 推荐增广路。一组边的集合, 要求从一个用户节点出发, 依次经过非匹配边、匹配边、...、非匹配边, 并且最后的项目节点项目容量不为 0, 则这样的路径称为推荐增广路。推荐系统二分图如图 1 所示。假设实线表示匹配边, 虚线表示非匹配边。  $u_3 \rightarrow i_1$ 、 $u_1 \rightarrow i_2$  为非匹配边,  $i_1 \rightarrow u_1$  为匹配边。假设此刻项目  $i_2$  的项目容量不为 0, 则  $u_3 \rightarrow i_1 \rightarrow u_1 \rightarrow i_2$  为一条推荐增广路。

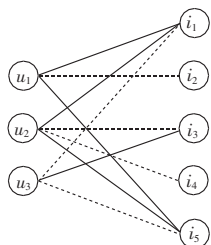


图1 推荐系统二分图

### 1.2 算法思想

推荐的过程通常分为预测用户未评分项目的评分和生成 top- $N$  推荐列表两个阶段<sup>[10]</sup>。因此, 提高推荐多样性往往有两种方法: a) 在预测评分时通过增加干预提高多样性; b) 在推荐生成阶段进行优化, 寻找一个最佳的推荐组合。本文采用方法 b), 提出 AD-Improved 算法。当用户推荐列表 top- $N$  未满足, 然而候选推荐列表中的项目均达到推荐次数限制时, 若将  $Cu$  中  $R^*(u, i)$  最高的项目添加到该用户  $Tu$  中, 则有利于提高推荐的准确率, 而不利于推荐的总体多样性; 若将系统中一个  $IC$  值

较大的冷门项目推荐给用户, 则有利于推荐的总体多样性, 而不利推荐的准确率。本文算法寻找从该用户节点开始的推荐增广路, 并置换增广路中的匹配边与非匹配边。以图 1 为例, 假设向每位用户推荐两个项目, 项目容量  $IC$  为 2。此刻  $u_3$  用户推荐列表未满足, 且候选推荐列表中项目  $i_1$  与  $i_5$  的  $IC$  值均为 0。AD-Improved 算法寻找到  $u_3 \rightarrow i_1 \rightarrow u_1 \rightarrow i_2$  增广路。将  $i_1$  推荐给  $u_3$ , 将  $i_2$  推荐给  $u_1$ 。由此实现了  $u_3$  用户的推荐, 并且增加了冷门项目  $i_2$  的推荐。将  $i_2$  推荐给  $u_1$  虽然可能会有损推荐准确率, 然而由于设置了  $R^*(u_1, i_2) > Tr$ ,  $u_1$  用户对  $i_2$  项目仍然具有较高的预测评分。因此, 在增强总体多样性的同时, 算法仍然具有较好的准确率。另外, 该算法可与其他任何具有评分预测步骤的推荐算法相结合, 以提高其推荐总体多样性。AD-Improved 算法可分为下列过程: 生成预测评分矩阵; 构建推荐二分图; 生成多样化的推荐列表。

### 1.3 生成预测评分矩阵

User-based CF、Item-based CF、Matrix Factorization CF 等现有推荐算法在向用户进行物品推荐时, 往往首先需要预测用户对未评分项目的评分, 并对每位用户选取评分最高的  $N$  个项目构建 top- $N$  推荐列表。本文算法利用现有推荐算法生成预测评分矩阵  $R^*$ 。

### 1.4 构建推荐二分图

构建推荐二分图过程如下: a) 设置评分阈值  $Tr$ , 并基于预测评分矩阵  $R^*$ , 获得用户候选推荐列表  $Cu$ ; b) 以非匹配边连接用户  $u$  与其  $Cu$  中的项目, 即为系统初始状态的推荐二分图。

### 1.5 生成推荐列表

向用户生成推荐列表, 共包括以下三种情况:

a) 用户  $Cu$  中存在  $IC$  不为 0 的项目, 为了保证推荐准确率, 选择预测评分值最高的项目  $i$  推荐给用户, 即将项目  $i$  添加到  $Tu$  中, 并从  $Cu$  中移除项目  $i$ 。同时, 由于项目  $i$  进行了一次推荐, 需要对项目  $i$  的项目容量进行更新:  $IC_i - 1$ 。最后, 二分图中用户  $u$  与项目  $i$  间的边由非匹配边转变为匹配边。

b) 用户  $Cu$  中所有项目  $IC$  均为 0 时, 从该用户节点出发采用广度优先搜索策略可获得至少一条推荐增广路。推荐增广路起始于用户节点, 终止于项目节点。推荐增广路  $p = v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_{2n-1} \rightarrow v_{2n}$ 。  $v_{\text{odd}}$  为用户节点,  $v_{\text{even}}$  为项目节点。因此  $v_{2k-1} \rightarrow v_{2k}$  为非匹配边,  $v_{2k} \rightarrow v_{2k+1}$  为匹配边。将该推荐增广路  $v_{2k-1} \rightarrow v_{2k}$  转变为匹配边,  $v_{2k} \rightarrow v_{2k+1}$  转变为非匹配边。根据边的转换, 修改相应用户的  $Tu$  与  $Cu$ , 将  $v_{2k}$  代表的项目添加到  $v_{2k-1}$  的用户  $Tu$  中, 将  $v_{2k}$  代表的项目从  $v_{2k+1}$  的用户  $Cu$  中移除。最后对推荐增广路  $v_{2n}$  所代表项目的项目容量进行减 1 操作:  $IC_{v_{2n}} - 1$ 。

c) 用户候选推荐列表中的项目容量均为 0, 且不存在从该用户节点出发的推荐增广路。此刻无法提高系统总体多样性, 为了保证算法准确率, 将候选推荐列表中评分最高的项目添加到用户推荐列表中, 不改变项目的项目容量值。

生成推荐列表过程如下:

```

输入:  $R^*$ ,  $Tr$ ,  $IC$ 。
输出: top- $N$  推荐列表。
 $Cu = \{item | item \in I, u \in U \text{ and } R^*(u, item) > Tr\}$ 
 $E = \{(v_u, v_{item}) | u \in U, item \in I \text{ and } (u, item) \in Cu\}$ 
 $G = (U, I, E)$ 
while top- $N$  not full
  for each user  $u \in U$ 
    if  $\forall i \in IC_{item} \neq 0$  and  $item \in Cu$ 
      sort items  $\{items \in Cu \text{ and } IC_{item} \neq 0\}$ 
      add most highly predicted item  $i$  to  $Tu$ 
      remove item  $i$  from  $Cu$ 
       $IC_i - 1$ 

```

```

change( $v_u, v_i$ ) to mathing path
end if
if not  $\exists IC_{item} \neq 0$  and  $item \in Cu$ , but  $\exists$  a augmentPath
    remove  $item_{2k}$  from  $Cu_{2k+1}$ 
    add  $item_{2k}$  to  $Tu_{2k-1}$ 
    change  $v_{2k-1} \rightarrow v_{2k}$  to mathing path
    change  $v_{2k} \rightarrow v_{2k+1}$  to unmathing path
     $IC_{2n} - 1$ 
end if
if not  $\exists IC_{item} \neq 0$  and  $item \in Cu$ , also not  $\exists$  a augmentPath
    sort items  $\{ items \in Cu \}$ 
    add most highly predicted item  $i$  to  $Tu$ 
    remove item  $i$  from  $Cu$ 
end if
end for
end while

```

## 1.6 算法复杂度分析

对于由  $n$  位用户,  $m$  个项目构成的系统进行算法复杂度分析。假设推荐列表长度为  $N$ 。以最坏情况进行考虑, 假设用户的每次推荐都需要进行寻找推荐增广路, 增广路复杂度为  $O(m \times n)$ 。系统有  $n$  位用户且需要重复  $N$  次, 复杂度为  $O(N \times m^2 \times n)$ 。由于  $N$  通常为 5 或者 10 的常数, 相对于系统中用户数量与项目数量而言可以忽略, 本文算法的时间复杂度为  $O(m^2 \times n)$ 。

## 2 实验

本文采用推荐领域广泛使用的 MovieLens-1M 数据集 (http://www.grouplens.org 提供) 作为测试数据进行实验。为了更好地与文献 [10] Expectation 算法进行对比, 采用同样的方法对数据进行预处理, 处理后的数据集包含 2 830 位用户对 1 919 个项目的 775 176 项评分, 预处理后数据的稀疏性为 14.27%, 并使用目前最为流行的三种协同过滤技术 (User-based CF、Item-based CF、Matrix Factorization CF) 预测用户对未评分项目的评分, 从而获取候选推荐列表。User-based CF、Item-based CF 中, 选取 50 个相似度最高的邻居作为最近邻; Matrix Factorization CF 中, 使用 5-fold 交叉衍生法确定常规化常数  $\lambda$  为 0.043, 潜在因子数为 15, 迭代次数 30 次。通过对 MovieLens-1M 数据进行分析发现, 70% 的项目被评价次数低于 50 次。针对该数据集, 设置项目容量  $IC$  为 50。

### 2.1 评价指标

本文使用 Adomavicius 等人 [9] 提出的 prediction-in-top- $N$  方法对推荐准确率进行度量。该算法在 Netflix 以及 MovieLens 数据集上进行不同的推荐算法和交叉验证实验, 预测准确度与平均预测评分的 person correlation 分别为 0.974 和 0.966。因此用平均预测评分可以有效地对推荐准确率进行估算。推荐准确率公式为

$$\text{prediction} = \frac{\sum_{u \in U} \sum_{i \in Tu} r_{u,i}^*}{\sum_{u \in U} |Tu|} \quad (1)$$

针对推荐总体多样性度量, 本文使用覆盖率 coverage、Gini-diversity 以及 entropy 三种指标, 公式如下:

$$\text{coverage} = |\bigcup_{u \in U} Tu| \quad (2)$$

$$\text{Gini-diversity} = - \sum_{j=1}^N (2j - n - 1) \frac{\text{rec}(i_j)}{|U| \times N} \quad (3)$$

$$\text{entropy} = - \sum_{i=1}^N \left( \frac{\text{rec}(i)}{|U| \times N} \right) \ln \left( \frac{\text{rec}(i)}{|U| \times N} \right) \quad (4)$$

coverage 值越大, 则系统覆盖的项目种类数越多, 一定程度上证明系统的总体多样性越好。Gini-diversity 与 entropy 从定量角度对推荐总体多样性进行度量。其中, Gini-diversity 值

越小, 多样性越好; entropy 值越大, 多样性越好。式中:  $\text{rec}(i)$  为项目  $i$  被推荐给不同用户的次数;  $i_j$  是按照物品流行程度由小到大排序中的第  $j$  个物品。

### 2.2 实验结果与分析

图 2 展示了不同评分阈值  $Tr = (3.5, 3.6, \dots, 5)$  下, 本文提出 AD-Improved 算法与文献 [10] Expectation 算法分别基于 User-based CF、Item-based CF、Matrix Factorization CF 进行 top-5 推荐在准确率与多样性指标下的实验对比结果。

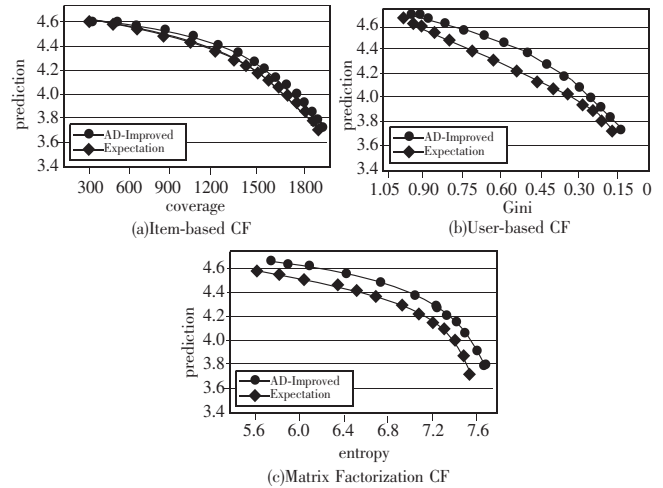


图 2 本文算法与 Expectation 算法基于不同现有推荐算法的准确率与多样性对比

从图 2(a) 可以看出, 基于 Item-based CF 的本文算法可灵活控制覆盖率与准确率的折中, 在阈值  $Tr = 4.5$  时, 准确率降低不超过 5%, 然而覆盖率提高近 4 倍, 与 Expectation 算法对比, 本文算法在准确率与覆盖率方面同样具有少量优势; 从 (b) 可以看出, 与 Expectation 算法相比, 在同一阈值的情况下, 基于 Item-based CF 的本文算法的 Gini 系数相对更小, 多样性方面具有微弱优势, 在准确率方面本文算法表现良好, 在阈值  $Tr \in [4, 4.6]$  时, 推荐准确率提高 3% 左右; 从 (c) 可以看出, 基于 Matrix Factorization CF 的本文算法在准确率与多样性方面同样要优于 Expectation 算法。为了更好地与 Expectation 算法进行对比, 同时展示准确率与多样性间的关系, 图 3 给出了在进行 top-5 推荐时, 通过准确率的调整 (通过评分阈值调整实现), 算法多样性方面的表现。

	User-based CF	Item-based CF	MF
standard	264 accuracy = 4.47	293 accuracy = 4.60	516 accuracy = 4.63
accuracy level: standard-0.01			
Expectation	480.8 (82.1%)	520.2 (77.5%)	657.9 (27.5%)
AD-Improved	513.2 (94.4%)	577.1 (97.0%)	667.7 (29.4%)
accuracy level: standard-0.05			
Expectation	734.2 (178.1%)	752.6 (156.9%)	816.8 (58.3%)
AD-Improved	773.0 (192.8%)	823.9 (181.2%)	890.1 (72.5%)
accuracy level: standard-0.1			
Expectation	890.5 (237.3%)	894.2 (205.1%)	1 050.0 (103.5%)
AD-Improved	998.9 (278.4%)	1 049.5 (258.2%)	1 150.2 (122.9%)

图 3 给定准确率条件下多样性提高情况

图 3 为传统 User-based CF、Item-based CF、Matrix Factorization CF 算法推荐准确率与多样性情况。通过图 3 可以了解到在分别损失准确率 0.01、0.05、0.1 的情况下, 本文算法在多样性方面具体的提高情况。即便是准确率仅仅损失 0.01, 基于 User-based CF 的本文算法与基于 Item-based CF 的本文算法的多样性仍然提高了 90% 以上。此外, 与 Expectation 算法相比, 本文算法在多样性提高方面表现更好。 (下转第 1667 页)

- [7] Newman M E J, Girvan M. Finding and evaluating community structure in networks[J]. *Physical Review E*, 2004, 69(2): 026113.
- [8] Newman M E J. Fast algorithm for detecting community structure in networks[J]. *Physical Review E*, 2004, 69(6): 066133.
- [9] Radicchi F, Castellano C, Cecconi F, *et al.* Defining and identifying communities in networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101(9): 2658-2663.
- [10] Gregory S. Local betweenness for finding communities in networks, Technical Report CSTR-08-004 [R]. Bristol: University of Bristol, 2008.
- [11] 朱小虎, 宋文军, 王崇骏, 等. 用于社团发现的 Girvan-Newman 改进算法[J]. *计算机科学与探索*, 2010, 4(12): 1101-1108.
- [12] 徐杨, 蒙祖强. 基于 GN 算法的微博社区识别方法[J]. *广西大学学报: 自然科学版*, 2013, 38(6): 1413-1417.
- [13] Yang Liang, Cao Xiaochun, Jin Di, *et al.* A unified semi-supervised community detection framework using latent space graph regularization [J]. *IEEE Trans on Cybernetics*, 2015, 45(11): 2585-2598.
- [14] Liu Zhiyuan, Li Peng, Zheng Yabin, *et al.* Community detection by affinity propagation [EB/OL]. 2008. [http://nlp.csai.tsinghua.edu.cn/~lzy/techreports/TR001\\_thunlp\\_community\\_detection.pdf](http://nlp.csai.tsinghua.edu.cn/~lzy/techreports/TR001_thunlp_community_detection.pdf).
- [15] Ver Steeg G, Galstyan A, Allahverdyan A E. Statistical mechanics of semi-supervised clustering in sparse graphs[J]. *Journal of Statistical Mechanics Theory and Experiment*, 2011(8): DOI: 10.1088/1742-5468/2011/08/P08009.
- [16] Ma Xiaoke, Gao Lin, Yong Xuerong, *et al.* Semi-supervised clustering algorithm for community detection in complex networks [J]. *Physica A: Statistical Mechanics and Its Applications*, 2010, 389(1): 187-197.
- [17] Zhang Zhongyuan. Community structure detection in complex networks with partial background information[J]. *Europhysics Letters*, 2013, 101(4): DOI:10.1209/0295-5075/101/48005.
- [18] Zhang Zhongyuan, Sun Kaidi, Wang Siqi. Enhanced community structure detection in complex networks with partial background information [J]. *Scientific Reports*, 2013, 3(11): DOI:10.1038/srep3241.
- [19] Karthik S, Aggarwal C C, Jaideep S, *et al.* Community detection with prior knowledge [C]//Proc of the SIAM International Conference on Data Mining. Philadelphia, PA: SIAM, 2013.
- [20] Cheng Jianjun, Leng Mingwei, Li Longjie, *et al.* Active semi-supervised community detection based on must-link and cannot-link constraints [J]. *Plos One*, 2014, 9(10): e110088.
- [21] Liu Dong, Bai Hongyu, Li Huijia, *et al.* Semi-supervised community detection using label propagation [J]. *International Journal of Modern Physics B*, 2014, 28(29): 1450208.
- [22] Kingma D, Rezende D, Mohamed S, *et al.* Semi-supervised learning with deep generative models [C]//Advances in Neural Information Processing Systems. 2014.
- [23] 姜雅文. 复杂网络社区发现若干问题研究 [D]. 北京: 北京交通大学, 2014.
- [24] Chai Bianfang, Yu Jian, Jia Caiyan, *et al.* Combining a popularity-productivity stochastic block model with a discriminative-content model for general structure detection [J]. *Physical Review E*, 2013, 88(1): 012807.
- [25] 柴变芳, 于剑, 贾彩燕, 等. 一种基于随机块模型的快速广义社区发现算法[J]. *软件学报*, 2013, 24(11): 2699-2709.
- [26] 柴变芳, 贾彩燕, 于剑. 基于概率模型的大规模网络结构发现方法[J]. *软件学报*, 2014, 25(12): 2753-2766.
- [27] 柴变芳, 赵晓鹏, 贾彩燕, 等. 内容网络广义社区发现有效算法[J]. *计算机科学与探索*, 2014, 8(9): 1076-1084.
- [28] 柴变芳, 赵晓鹏, 贾彩燕, 等. 大规模网络的三角形模体社区发现模型[J]. *南京大学学报: 自然科学版*, 2014, 50(4): 466-473.
- [29] Guerra L, Bieza C, Robles V. Semi-supervised projected model-based clustering [J]. *Data Mining and Knowledge Discovery*, 2014, 28(4): 882-917.

(上接第 1630 页)

### 3 结束语

推荐多样性正日益成为评价推荐系统性能的重要指标。针对现有推荐算法缺少对推荐总体多样性考虑现象, 提出一种基于二分图网络的推荐多样性增强算法。通过对现有推荐算法进行二次优化, 与现有算法进行实验对比, 结果表明本文算法能有效提高系统总体多样性且保持较高推荐准确率。本文算法的不足之处在于推荐增广路的选取上缺少对算法准确率的考虑, 没有考虑推荐的个体多样性。如何同时提高系统个体多样性与总体多样性, 将是下一步的研究方向。

#### 参考文献:

- [1] Hill W, Stead L, Rosenstein M, *et al.* Recommending and evaluating choices in a virtual community of use [C]//Proc of ACM SIGCHI Conference on Human Factors in Computing Systems. New York: ACM Press, 1995: 194-201.
- [2] Resnick P, Varian H R. Recommender systems[J]. *Communications of the ACM*, 1997, 40(3): 56-58.
- [3] Kim J E, Kim H G. Music recommendation method with respect to message service: US, US8410347 [P]. 2013.
- [4] 陈克寒, 韩盼盼, 吴健. 基于用户聚类的异构社交网络推荐算法[J]. *计算机学报*, 2013, 36(2): 349-359.
- [5] 安维, 刘启华, 张李义. 个性化推荐系统的多样性研究进展[J]. *图书情报工作*, 2013, 57(20): 127-135.
- [6] Ziegler C N, Mcnee S M, Konstan J A, *et al.* Improving recommendation lists through topic diversification [M]//A History of the Comstock Silver Lode & Mines. [S. l.]: Promontory Press, 1974: 22-32.
- [7] Niemann K, Wolpers M. A new collaborative filtering approach for increasing the aggregate diversity of recommender systems [C]//Proc of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2013: 955-963.
- [8] Adomavicius G, Kwon Y O. Maximizing aggregate recommendation diversity: a graph-theoretic approach [C]//Proc of the 1st International Workshop on Novelty and Diversity in Recommender Systems. 2011: 3-10.
- [9] Adomavicius G, Kwon Y O. Improving aggregate recommendation diversity using ranking-based techniques [J]. *IEEE Trans on Knowledge & Data Engineering*, 2012, 24(5): 896-911.
- [10] 刘慧婷, 岳可诚. 可提高多样性的基于推荐期望的 top-N 推荐方法[J]. *计算机科学*, 2014, 41(7): 270-274.