

# 协同过滤推荐算法在豆瓣网络数据上的研究<sup>①</sup>

马晓迪, 宣琦, 张哲, 傅晨波, 董辉

(浙江工业大学 信息工程学院, 杭州 310023)

**摘要:** 在豆瓣网络数据上对传统的协同过滤推荐算法进行改进, 分别考虑最近邻和有向相似度的作用, 对图书、电影和音乐收藏列表进行个性化推荐. 推荐的结果在准确度、多样性和新奇性三种被广泛使用在衡量推荐算法效果的指标上进行比较和分析. 结果表明, 相比传统协同过滤推荐算法, 两种改进算法均能够保证多样性和新奇性, 同时最近邻算法可有效降低算法复杂度, 而有向相似度算法则具有更高的推荐准确度.

**关键词:** 复杂网络; 推荐算法; 协同过滤; 最近邻; 有向相似度

## Research of Collaborative Filtering Recommendation Algorithm on Douban Network Data

MA Xiao-Di, XUAN Qi, ZHANG Zhe, FU Chen-Bo, DONG Hui

(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310032, China)

**Abstract:** This paper improved the collaborative filtering recommendation algorithm by considering the nearest neighbor and directed similarity in Douban network data. Then, the improved algorithms were used to recommend books, movies and music for Douban users. The recommended results are carefully compared and analyzed in terms of three well-know indicators including accuracy, diversity and novelty. It is shown that the nearest neighbor algorithm has much lower computational complexity and the directed similarity algorithm obtains higher accuracy, while all these three algorithms have similar diversity and novelty of the recommended results, by comparing with the traditional collaborative filtering recommendation algorithm.

**Key words:** complex networks; recommendation algorithm; collaborative filtering; nearest neighbor; directed similarity

随着互联网技术的高速发展, 网络中的信息量呈几何级数增长. 快速增长的信息远远超过了人们的处理能力, 从而降低了人们对其的使用效率, 并形成信息过载<sup>[1]</sup>. 个性化推荐作为目前解决信息过载的有效途径之一, 得到了国内外许多研究人员的关注<sup>[2,3]</sup>. 截止目前, 研究者已提出了各种算法, 如基于内容的推荐算法<sup>[4-6]</sup>、基于知识的推荐算法<sup>[7]</sup>、基于网络结构的推荐算法<sup>[8,9]</sup>、协同过滤算法<sup>[10,11]</sup>、混合推荐算法<sup>[12,13]</sup>等. 最近, 数据挖掘以及机器学习领域中的方法也被引入到了推荐算法中, 以期待进一步提升算法性能<sup>[14]</sup>. 其中, 协同过滤推荐 (Collaborative Filtering Recommendation) 是目前应用最为广泛且最成功的个性化推荐算法. 其基本思想是利用用户对条目的评价

历史计算用户间相似度, 借鉴与目标用户相似度较高的邻居对其它条目的评价, 来针对目标用户进行推荐. 早在 1979 年, Grundy 书籍推荐系统<sup>[15]</sup>就已建立用户兴趣爱好模型, 用于向用户推荐书籍, 是公认的第一个真正意义上投入实际应用的协同过滤推荐系统. 此外, 类似的系统还包括 Tapestry 邮件处理系统<sup>[16]</sup>、Grouplens 新闻推荐系统<sup>[17]</sup>、Ringo 音乐推荐系统<sup>[18]</sup>等. 值得一提的是 Amazon 的协同过滤推荐系统, 通过为用户推荐可能感兴趣的商品, 提高了 35% 的销售额<sup>[19,20]</sup>. 近年来国内外研究人员针对协同过滤推荐也提出了很多改进策略, 如针对用户评分稀疏性和对推荐实时性的要求, 李聪等人<sup>[21]</sup>提出了基于邻域最近邻的协同过滤, 并对非目标用户的推荐能力进行评估区分, 提高

<sup>①</sup>基金项目: 国家自然科学基金青年基金(61004097); 国家自然科学基金(61273232)

收稿时间: 2013-12-12; 收到修改稿时间: 2014-01-17

了算法的实时性. 刘建国等人<sup>[22]</sup>研究发现用户相似性的方向<sup>[23]</sup>可对推荐结果有一定的影响. 但他们或是利用模拟网络数据进行计算, 或是利用网络上著名的 Movielens<sup>[24]</sup>数据集进行验证, 而并没有将最近邻算法, 有向相似度算法, 以及传统的协同过滤算法进行横向比较. 鉴于这两种推荐策略从不同方向上对协同过滤算法进行优化, 故有必要对它们进行比较研究.

本文对豆瓣网络数据进行分析, 在此基础上对传统的协调过滤算法进行优化. 我们分别采用全局相似度策略, 最近邻用户相似度策略, 并考虑相似度方向性, 比较分析不同策略下的算法在不同评价指标上的推荐效果. 算法评价指标包括精度、召回率, 以及推荐结果的多样性和新颖性等. 本文的结构如下: 第一节介绍豆瓣网络实际数据预处理, 第二节描述推荐算法, 第三节提出评价指标, 第四节对推荐结果进行分析, 最后一节给出总结和讨论. 本文对豆瓣网络数据进行分析, 在此基础上对传统的协调过滤算法进行优化. 我们分别采用全局相似度策略, 最近邻用户相似度策略, 并考虑相似度方向性, 比较分析不同策略下的算法在不同评价指标上的推荐效果. 算法评价指标包括精度、召回率, 以及推荐结果的多样性和新颖性等. 本文的结构如下: 第一节介绍豆瓣网络实际数据预处理, 第二节描述推荐算法, 第三节提出评价指标, 第四节对推荐结果进行分析, 最后一节给出总结和讨论.

## 1 数据预处理

我们采集了 2008 年 9 月 1 日至 2009 年 5 月 1 日的豆瓣网用户在图书、电影和音乐领域的收藏条目数据, 用于测试推荐算法. 该数据包含用户在豆瓣网上所收藏的不同类型条目 ID 和收藏时间. 其中 2008 年 9 月 1 日至 2009 年 3 月 1 日期间的数据作为训练集用于构建算法, 2009 年 3 月 2 日至 5 月 1 日的数据作为测试集用于测试算法, 定义  $B_i$  为用户  $i$  在训练集的收藏条目集合,  $|B_i|$  为训练集收藏条目数, 定义  $D_i$  为用户在测试集的收藏条目集合,  $|D_i|$  为测试集收藏条目数. 本文中, 我们选取满足  $20 \leq |B_i| \leq 200$  以及  $|D_i| > 0$  的用户进行研究, 理由如下: 对于训练集中收藏过少的用户, 他与其他用户之间的相似性很难被衡量, 而对于训练集中短期内收藏过多的疯狂收藏者, 其真正的收藏爱好同样也很难鉴定, 对于此类用户进行盲目推荐将极有可能降低他们的推荐体验, 故本文不考虑此类

用户符合实际的推荐目的. 具体的数据格式举例如表 1 所示, 在所研究的图书网络中, ID29 的用户在 2008 年 9 月 1 日至 2009 年 3 月 1 日期间收藏了 20 本图书, 即  $|B_{29}|=20$ , 在 2009 年 3 月 2 日至 5 月 1 日期间收藏了 2 本图书, 即  $|D_{29}|=2$ . 符合上文对数据的选取条件.

表 1 图书网络部分数据举例

用户 ID 编号	物品 ID 编号	收藏时间
29	2846	2009-4-1
29	2847	2009-3-21
29	2848	2009-1-30
29	2849	2009-1-20
29	479	2009-1-17
29	2457	2009-1-16
29	2349	2009-1-8
29	2850	2008-12-22
29	2341	2008-12-17
29	275	2008-12-6
29	2343	2008-11-9
29	2346	2008-11-9
29	2342	2008-11-9
29	523	2008-10-28
29	2405	2008-10-27
29	2406	2008-10-22
29	2851	2008-10-20
29	2852	2008-10-12
29	1570	2008-10-1
29	2411	2008-9-15
29	2853	2008-9-13
29	2854	2008-9-4

## 2 算法描述

依照协同过滤算法出发点的不同, 传统的协同过滤算法一般可分为基于用户-用户关系的协同过滤算法和基于条目-条目关系的协同过滤算法. 前者通过分析不同用户对条目的评分得到用户间的相似度, 然后基于用户之间的相似度进行推荐; 而后者则通过分析用户对条目的评分或条目属性得到条目间的相似度然后基于条目之间的相似度做推荐. 研究发现, 在用户数量远大于条目数量的数据上, 基于用户-用户的协同过滤算法效果较好, 反之, 这一算法的有效性就不一定能得到保障<sup>[25]</sup>. 由于豆瓣数据集中, 用户数量远大于条目数量, 而且数据集仅包含用户的收藏行为记录, 而没有条目的内容分类, 用户对条目的评分也仅表现

于是否收藏,因此我们采用基于用户-用户关系的协同过滤算法.该算法基于以下假设:若用户对某些条目的评分相似,则他们对于其他条目的评分也会相似.基于训练集的定义,用户*i*与用户*j*的相似度 $S_{ij}$ 定义为

$$S_{ij} = \frac{|B_i \cap B_j|}{|B_i \cup B_j|}, \quad (1)$$

对于用户*i*而言,训练集中未添加条目*k*的推荐得分记为 $Score_i(k)$ ,定义如下:

$$Score_i(k) = \frac{\sum_{j \in \pi_i} S_{ij} \delta_j(k)}{\sum_{j \in \pi_i} S_{ij}}, \quad (2)$$

其中, $\pi_i$ 为相似度网络中用户*i*的邻居,当且仅当条目*k*已在用户*j*训练集收藏列表中时, $\delta_j(k)=1$ ,否则为0.

然而,通过对豆瓣网络数据的分析,我们发现由式(1)得到的相似度网络存在不足:相似度网络连边密度较大,而用户间相似度平均值普遍较低,即相似度网络中目标用户有很多邻居,但用户与大部分邻居的收藏记录相似度很低,往往只有少数几个条目是共同收藏的,这样绝大部分邻居只能提供与目标用户相关的微弱信息,难以对推荐结果产生影响.由于实际数据中通常伴随一定强度的噪声,因此,同时考虑大量邻居提供的微弱有效信息反而有可能降低协同推荐算法的精度.同时连边密度过高也会导致计算时间复杂度的急剧上升,从而降低推荐效率.

为了扩大真正相似用户在推荐过程中的作用,减少噪声的干扰,提升实时推荐效率,我们对传统的协同过滤算法进行两个方向上的改进.第一个改进是裁剪原有相似度网络:对每个用户推荐时仅计算相似度最大的 $\theta$ 个邻居的评分影响,不考虑其他相似度较小的邻居.这里有两点需要作出说明:首先,如果 $S_{ij}$ 不在*i*的前 $\theta$ 个最大相似度内,却在*j*的前 $\theta$ 个最大相似度内,则在对*i*进行推荐时不考虑邻居*j*的评分影响,而在对*j*进行推荐时需考虑邻居*i*的评分影响;其次,如果相似度网络中目标用户的邻居数小于 $\theta$ ,则在计算评分时考虑其所有的邻居.第二个改进是在相似度网络中考虑方向性,即认为在收藏条目不同的两用户间,相似度具有方向性,这样可以削弱收藏条目过多的用户的影响力.对于目标用户*i*,其有向相似度定义为

$$S_{ij} = \frac{|B_i \cap B_j|}{|B_j|}. \quad (3)$$

在第四节中,对于第一节所选取的训练集,我们将分别执行以下三种协同推荐算法:结合式(1)和式(2)的考虑全局相似度网络的传统协同过滤推荐算法;结合式(1)和式(2)的考虑最近邻原则的局部相似度网络算法;以及结合式(3)和式(2)的考虑有向相似度算法.同时,我们也将在此实际的豆瓣网络数据上执行两种额外的推荐算法:随机推荐算法和全局度排序算法,作为基准.然后,我们将通过对五种推荐算法获得的推荐结果进行比较,说明最近邻和方向性策略在为豆瓣用户进行推荐时的作用.

### 3 评价指标

评价推荐系统效果的指标有很多种,如准确度<sup>[26,27]</sup>、覆盖率<sup>[28,29]</sup>、产品流行性<sup>[30]</sup>、多样性<sup>[9]</sup>和新颖性<sup>[2,30]</sup>等.目前就如何选择恰当的评价指标来衡量推荐算法的表现在学术界尚未达成共识.鉴于此,本文将从准确度、多样性和新颖性等多个方面对算法进行评测,以期获得较为全面的算法效果比较.

#### 3.1 准确度

准确度考虑测试集中用户收藏条目在推荐列表前*L*个位置里的数目<sup>[23]</sup>.基于准确度通行的度量标准是精度和召回率.对于目标用户*i*,推荐的精度和召回率 $P_i(L)$ , $R_i(L)$ 定义如下<sup>[23,26-28]</sup>

$$P_i(L) = \frac{d_i(L)}{L}, R_i(L) = \frac{d_i(L)}{|D_i|}, \quad (4)$$

这里 $d_i(L)$ 表示测试集中用户的收藏条目在推荐列表前*L*位中的数目, $|D_i|$ 是测试集中的收藏条目总数.将所有在测试集中至少有一个收藏条目的用户的精度和召回率平均化,我们可以得到平均精度和平均召回率,分别记为 $P(L)$ 和 $R(L)$ .这些值可以与随机推荐系统的结果进行比较.通常而言,精度随增加而降低,而召回率*L*随增加而增加.

#### 3.2 多样性

多样性是表征所推荐条目之间的差异程度<sup>[30-32]</sup>.通常有两个层面来解释多样性:一是指算法针对不同用户返回不同结果的能力——称之为用户间多样性(即推荐列表间的多样性);另一个是衡量算法能够给每个个体用户提供多样性条目的程度——称之为用户

内多样性(即一个推荐列表内的多样性). 本文我们主要考虑的多样性是指用户间多样性. 用户间多样性可由海明距离来衡量, 给出用户  $i$  和  $j$ , 通过海明距离来衡量他们推荐列表中前  $L$  位的区别<sup>[9]</sup>

$$H_{ij}(L) = 1 - \frac{Q_{ij}(L)}{L}, \quad (5)$$

其中  $Q_{ij}(L)$  是两用户推荐列表前  $L$  位的公共产品数. 如果列表完全一致,  $H_{ij}(L)=0$ , 而如果列表完全不同则  $H_{ij}(L)=1$ . 对所有用户对平均化  $H_{ij}(L)$ , 可以获得平均海明距离  $H(L)$ . 该值越大, 表示系统给予的推荐越多样化.

### 3.3 新颖性

新颖性是指推荐条目与用户可能已知的条目相比差异的程度<sup>[30]</sup>. 量化算法的这一能力最简单的方式是衡量推荐产品的平均人气<sup>[31,32]</sup>, 即产品被多少用户所收藏, 定义如下:

$$D_{avg}(L) = \frac{1}{ML} \sum_{i=1}^M \sum_{\alpha \in O_k^i} k_{\alpha}, \quad (6)$$

其中  $M$  表示用户总量,  $L$  表示推荐列表长度,  $O_k^i$  表示用户  $i$  的推荐列表而  $k_{\alpha}$  表示条目的  $a$  度值(即人气), 低人气意味着高新奇度.

## 4 结果与分析

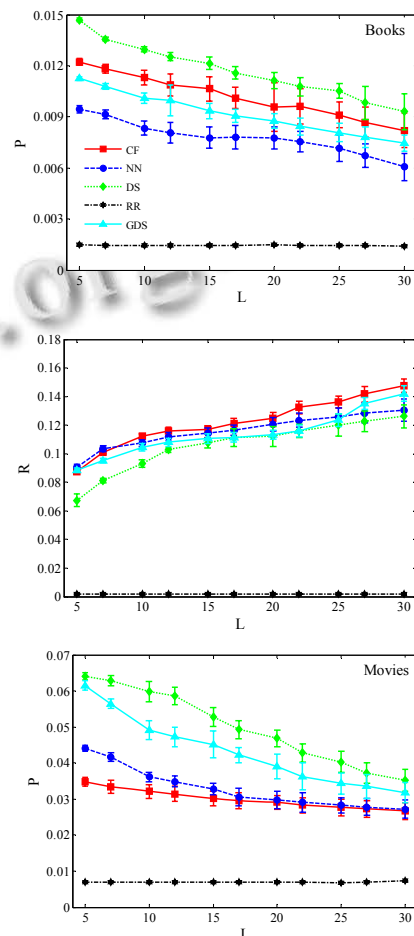
我们从豆瓣网络的训练集中每次随机选取 1000 个用户分别执行传统协同过滤算法(Collaborative Filtering, CF)、最近邻算法(Nearest Neighbor, NN)和有向相似算法(Directed Similarity, DS), 对每一种算法分别进行 50 次实验. 其中最近邻网络是选取与目标用户相似度最大的 5 个邻居与目标用户所形成的相似网络. 此外, 为了与以上这三种算法相比较, 本文采用两种非个性化推荐算法: 全局度排序算法(Global Degree Sorting, GDS)和随机推荐算法(Random Recommendation, RR)作为对照算法.

全局度排序算法的步骤是: 1)计算训练集中所有条目被用户收藏的数量(即条目的度); 2)将条目按照度值从大到小降序排列形成条目推荐列表, 排在列表首位的即是被最多用户收藏过的条目; 3)将条目推荐列表的前  $N$  项按度降序排列顺序推荐给用户, 其中  $N$  为推荐数量. 该算法的优点是时间复杂度极低, 因为条目推荐列表一旦生成就可以直接向所有用户返回推荐

结果, 但缺点也十分明显, 即完全没有进行个性化推荐, 所有用户得到的都是相同的按照度值降序排列的推荐列表. 随机推荐算法的步骤是: 1)将训练集中的所有条目随机选取其中  $N$  项条目形成推荐列表; 2)将推荐列表返回给用户. 可见不同用户间的推荐结果基本没有关联, 甚至对于同一用户, 每次的推荐结果都可能完全不同, 随机推荐的结果多样性可以得到保证, 但由于其没有利用任何有效的用户行为信息, 其精度通常都非常低. 以上这两种非个性化推荐算法作为经典的对比算法, 被广泛应用在与个性化推荐算法的对比和结果评价上<sup>[7]</sup>. 我们通过对推荐结果在准确度、多样性和新奇性三个指标上进行比较分析来衡量不同算法的推荐效果.

### 4.1 精度和召回率

分别针对豆瓣网络中收藏的书、电影和音乐, 对每一种推荐算法和评价指标, 我们记录 50 次实验的平均值和标准差. 在不同推荐列表长度下的算法精度和召回率, 如图 1 所示, 其中推荐列表长度取值为 5 至 30.



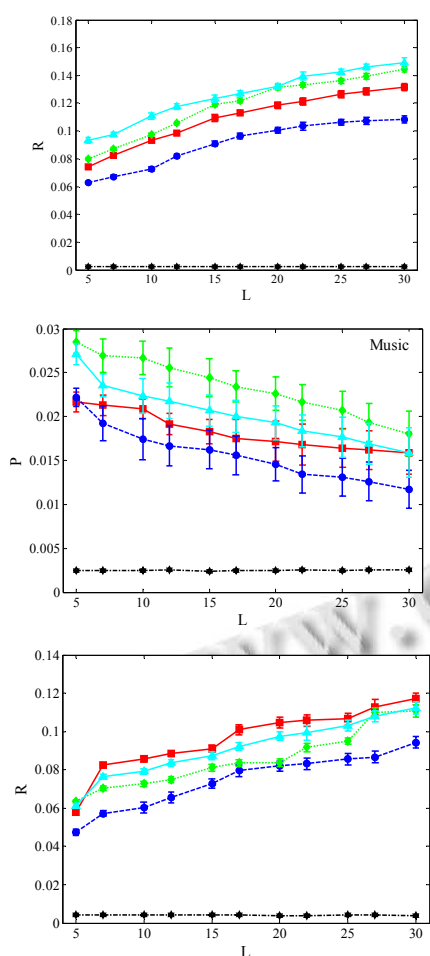


图 1 图书、电影、音乐网络下几种算法的精度和召回率，其中横坐标为推荐列表长度，纵坐标为算法精度，为算法召回率。

我们发现，在对豆瓣网络中图书、电影和音乐三

种收藏进行推荐的过程中，有向相似度算法(DS)(图中绿菱点线)在精度方面较全局度排序算法(GDS)(图中浅蓝三角点线)、传统协同过滤算法(CF)(图中红方点线)和最近邻算法(NN)(图中深蓝圆点线)均有很大的提高，但在时间复杂度上，因为有向相似度算法要计算用户间相互的相似度，其时间复杂度大大高于传统协同过滤算法(CF)。值得一提的是，在音乐网络中，随着推荐列表长度的增加，有向相似度算法(DS)的精度衰减速度似乎要快于其他算法。然而，在实际应用中，由于推荐列表的长度通常控制在 30 以内，故而从我们仍然可以认为有向相似度算法(DS)的精度普遍优于本文中的其他算法。对于最近邻算法(NN)，其精度仅在电影网络中略优于传统协同过滤算法(CF)，在其余两个网络中表现均较差。这可能是由于仅考虑最近邻相似度的算法在极大地减少了算法复杂度的同时，也忽略了网络中的一部分有效信息，造成了在去除信息噪声的同时，也会丢失部分有价值信。对于最近邻算法，如何选择最恰当的邻居数在提高算法速度的同时保证推荐精度是将来值得深入研究的方向之一。

#### 4.2 多样性

多样性的评判标准如前文所述，采用计算推荐内容的平均海明距离的方法，平均海明距离越大，多样性越强。本文中，针对 50 次实验，我们分别给出 50 次平均海明距离的平均值和标准差。全局度排序算法(GDS)向所有用户给出完全相同的推荐结果，故其平均海明距离数值上等于 0。其他三种推荐方法与随机推荐的海明距离比较如表 2 所示。

表 1 平均海明距离比较

推荐算法	CF	NN	DS	RR	
图书网络	L=10	0.924±0.005	0.945±0.008	0.914±0.006	0.971±0.002
	L=20	0.928±0.007	0.948±0.009	0.918±0.005	0.986±0.001
	L=30	0.932±0.004	0.955±0.008	0.927±0.006	0.990±0.001
电影网络	L=10	0.842±0.012	0.857±0.014	0.858±0.009	0.944±0.002
	L=20	0.850±0.009	0.860±0.011	0.866±0.011	0.971±0.003
	L=30	0.850±0.007	0.877±0.009	0.869±0.006	0.981±0.001
音乐网络	L=10	0.916±0.011	0.914±0.013	0.885±0.007	0.960±0.003
	L=20	0.922±0.009	0.925±0.014	0.894±0.007	0.984±0.002
	L=30	0.922±0.009	0.946±0.011	0.899±0.005	0.991±0.002

总体而言，随机推荐算法具有最好的多样性，最近邻算法(NN)次之，而有向相似度算法(DS)则相对较

差。即便如此，所有个性化推荐算法都具有不错的多样性，所有平均海明距离均大于 0.8。即对于任意两个

用户, 推荐列表中至少有 80% 的条目是彼此互不相同的。

有意思的是, 在图书和音乐网络中, 最近邻算法 (NN) 和有向相似性算法 (DS) 的平均海明距离均在 0.9 左右, 而在电影网络中, 推荐列表多样性则有所降低。用传统协同过滤 (CF) 算法也得出了类似的结果。分析网络特性后我们发现电影网络中大度值条目的度值比图书、音乐网络高很多, 即相比图书和音乐, 电影的流行性更强, 从而导致用户的兴趣点比较趋同。

### 4.3 新颖性

本文采用衡量推荐条目的平均人气来比较不同算法的新颖性。如图 2 所示, 与算法准确度评估类似, 推荐列表长度取值 5 至 30。一般认为, 推荐条目的平均人气越低, 其推荐系统就具有越好的新颖性。几种协同过滤算法在新颖性方面均明显好于全局度推荐算法。

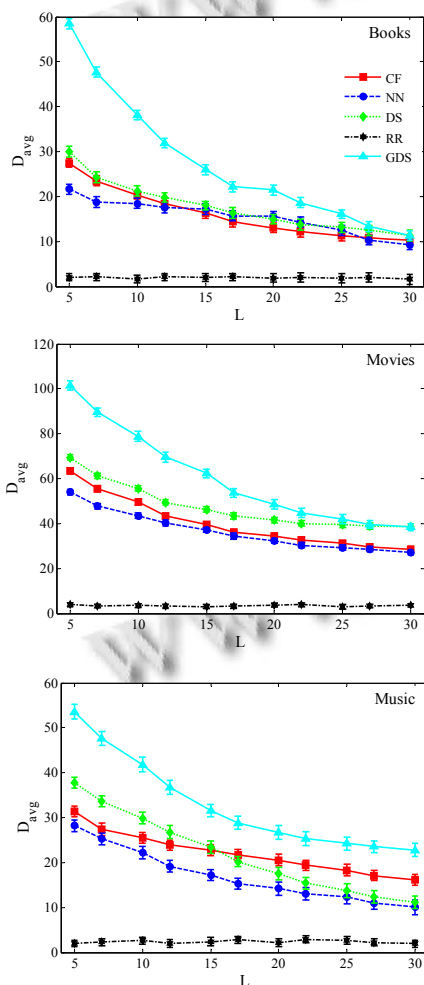


图 2 推荐条目平均人气比较(其中横坐标  $L$  为推荐列表长度, 纵坐标  $D_{avg}$  为推荐条目的平均人气)

## 5 结论与讨论

本文分析比较了传统协同过滤算法、最近邻算法和有向相似性算法以及常用的随机推荐和全局度推荐在豆瓣网络数据上的推荐效果。为了综合衡量算法的优劣, 我们从准确度、多样性和新颖性三个方面对推荐算法进行比较。在准确度指标上, 有向相似性算法的精度均好于其他两种协同过滤算法和全局度推荐算法。而最近邻算法则具有最低的计算复杂度, 可以很好地解决实时推荐的问题, 这也验证了李聪等人论文中的结论。在多样性指标上, 相比较完全没有个性化的全局度推荐, 三种协同过滤算法均表现了较好的推荐结果多样性。在新颖性指标上, 最近邻算法返回的推荐结果平均度较低, 可以为用户提供较为新奇的条目, 有向相似性算法表现略差, 尤其是在电影网络中, 推荐结果的平均度比较大。通过对比研究我们发现, 电影网络中大度值条目被大量用户所收藏, 用户选择具有较高的相似性, 因此在推荐打分中更易获得较高的分数, 出现在推荐列表的前端, 这一结果与上文 4.2 节多样性指标的讨论可相互印证。

协同过滤算法作为比较经典和主流的推荐算法已经发展了多年, 国内外研究人员也对其进行了各种改进, 衡量推荐算法优劣的评价指标有很多, 本文仅从准确度、多样性和新奇性三个指标来进行比较, 接下来的工作将会对算法在推荐结果覆盖率、用户满意度以及准确度的其他方面如推荐列表内的排序准确度、半衰期效用指标上的表现进行对比研究。最近邻算法因为算法复杂度较低可用于对于在线推荐要求比较高的系统中, 而有向相似性算法的推荐精度较高, 可以在数据规模比较小的网络中进行推荐。今后随着互联网科学的发展, 推荐算法必将会对电子商务、社交网络产生巨大的影响, 为互联网信息的过载提供行之有效的解决途径。

### 参考文献

- 1 Maes P. Agents that reduce work and information overload. *Communications of the ACM*, 1994, 37(7): 30–40.
- 2 Lü L, Medo M, Yeung CH, Zhang YC, Zhang ZK, Zhou T. Recommender systems. *Physics Reports*, 2012, 519(1): 1–49.
- 3 刘建国, 周涛, 汪秉宏. 个性化推荐系统的研究进展. *自然科学进展*, 2009, 19(1): 1–15.
- 4 Pazzani M, Billsus D. Learning and revising user profiles: The identification of interesting web sites. *Machine Learning*,

- 1997, 27(3): 313–331.
- 5 Mooney RJ, Bennett PN, Roy L. Book recommending using text categorization with extracted information. Proc. Recommender Systems Papers from 1998 Workshop, [Technical Report] WS-98-08. 1998.
  - 6 Chang YI, Shen JH, Chen TI. A data mining-based method for the incremental update of supporting personalized information filtering. *Journal of Information Science and Engineering*, 2008, 24(1): 129–142.
  - 7 张丙奇. 基于领域知识的个性化推荐算法研究. *计算机工程*, 2005, 21.
  - 8 Zhou T, Ren J, Medo M, Zhang YC. Bipartite network projection and personal recommendation. *Physical Review E*, 2007, 76(4): 046115.
  - 9 Zhou T, Jiang LL, Su RQ, Zhang YC. Effect of initial configuration on network-based recommendation. *EPL (Europhysics Letters)*, 2008, 81(5): 58004.
  - 10 Liu RR, Jia CX, Zhou T, Sun D, Wang BH. Personal recommendation via modified collaborative filtering. *Physica A: Statistical Mechanics and its Applications*, 2009, 388(4): 462–468.
  - 11 Balabanovi M, Shoham Y. Fab: content-based, collaborative recommendation. *Communications of the ACM*, 1997, 40(3): 66–72.
  - 12 Soboroff I, Nicholas C. Combining content and collaboration in text filtering. Proc. of the IJCAI. 1999, 99. 86–91.
  - 13 Yoshii K, Goto M, Komatani K, Ogata T. An efficient hybrid music recommender system using an incrementally trainable probabilistic generative model. *IEEE Trans. on Actions, Audio, Speech, and Language Processing*, 2008, 16(2): 435–447.
  - 14 Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. on Knowledge and Data Engineering*, 2005, 17(6): 734–749.
  - 15 Rich E. User modeling via stereotypes. *Cognitive Science*, 1979, 3(4): 329–354.
  - 16 Goldberg D, Nichols D, Oki BM, Terry D. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 1992, 35(12): 61–70.
  - 17 Konstan JA, Miller BN, Maltz D, Herlocker JL. GroupLens: applying collaborative filtering to Usenet news. *Communications of the ACM*, 1997, 40(3): 77–87.
  - 18 Shardanand U, Maes P. Social information filtering: algorithms for automating “word of mouth”. Proc. of the SIGCHI conference on Human factors in computing systems. ACM Press/Addison-Wesley Publishing Co., 1995. 210–217.
  - 19 Linden G, Smith B, York J. Amazon. com recommendations: Item-to-item collaborative filtering. *Internet Computing, IEEE*, 2003, 7(1): 76–80.
  - 20 张亮, 柏林森, 周涛. 基于跨电商行为的交叉推荐算法. *电子科技大学学报*, 2013, 42(1).
  - 21 李聪, 梁昌勇, 马丽. 基于领域最近邻的协同过滤推荐算法. *计算机研究与发展*, 2008, 45(9): 1532–1538.
  - 22 石珂瑞, 刘建国, 郭强, 冷瑞. 有向相似性对协同过滤推荐系统的影响研究. *复杂系统与复杂性科学*, 2012, 9(3): 46–49.
  - 23 Albert R, Jeong H, Barabási AL. Error and attack tolerance of complex networks. *Nature*, 2000, 406(6794): 378–382.
  - 24 Miller BN, Albert I, Lam SK, Konstan JA. MovieLens unplugged: experiences with an occasionally connected recommender system. Proc. of the 8th International Conference on Intelligent User Interfaces. ACM. 2003. 263–266.
  - 25 王国霞, 刘贺平. 个性化推荐系统综述. *计算机工程与应用*, 2012, 48(7): 66–76.
  - 26 刘建国, 周涛, 郭强, 汪秉宏. 个性化推荐系统评价方法综述. *复杂系统与复杂性科学*, 2009, 6(3): 1–10.
  - 27 Billsus D, Pazzani MJ. Learning collaborative information filters. *ICML*. 1998, 98: 46–54.
  - 28 Geyer-Schulz A, Hahsler M. Evaluation of recommender algorithms for an internet information broker based on simple association rules and on the repeat-buying theory. Proc. of WEBKDD. 2002. 100–114.
  - 29 Sarwar B, Karypis G, Konstan J, Riedl J. Analysis of recommendation algorithms for e-commerce. Proc. of the 2nd ACM Conference on Electronic Commerce. ACM. 2000. 158–167.
  - 30 McNee SM, Riedl J, Konstan JA. Being accurate is not enough: how accuracy metrics have hurt recommender systems. CHI’06 Extended Abstracts on Human Factors in Computing Systems. ACM. 2006. 1097–1101.
  - 31 Zhou T, Kuscsik Z, Liu JG, Medo M. Solving the apparent diversity-accuracy dilemma of recommender systems. Proc. of the National Academy of Sciences, 2010, 107(10): 4511–4515.
  - 32 Castells P, Vargas S, Wang J. Novelty and diversity metrics for recommender systems: choice, discovery and relevance. Proc. of International Workshop on Diversity in Document Retrieval (DDR). 2011. 29–37.