

信任社交网络中改进的物质扩散推荐算法

蔡永嘉, 李冠宇

(大连海事大学 信息科学技术学院, 辽宁 大连 116026)

摘 要: 传统物质扩散推荐算法多样性低且未考虑用户所处的社交网络信息和物品的受欢迎程度。为此, 在基于信任机制的社交网络中, 提出一种改进的物质扩散推荐算法。引入信任机制形成目标用户的最优邻居集, 模拟用户-物品二分网络, 根据用户信任度对物品初始资源进行分配。考虑物品双向扩散能力并结合物品流行度的可调参数, 实现资源再分配, 从而优化目标用户的推荐结果。在真实数据集上的实验结果表明, 该算法能在保证较高推荐准确率的同时, 增强推荐结果的多样性。

关键词: 物质扩散; 信任机制; 社交网络; 流行度; 二分网络

中文引用格式: 蔡永嘉, 李冠宇. 信任社交网络中改进的物质扩散推荐算法[J]. 计算机工程, 2019, 45(3): 182-187.

英文引用格式: CAI Yongjia, LI Guanyu. Improved mass diffusion recommendation algorithm in trusted social network[J]. Computer Engineering, 2019, 45(3): 182-187.

Improved Mass Diffusion Recommendation Algorithm in Trusted Social Network

CAI Yongjia, LI Guanyu

(Faculty of Information Science and Technology, Dalian Maritime University, Dalian, Liaoning 116026, China)

[Abstract] Traditional mass diffusion recommendation algorithm has low diversity and its neglect to user's social network information and object popularity. Therefore, this paper proposes a mass diffusion recommendation algorithm based on trust mechanism in social network. The trust mechanism is introduced to form an optimal neighbor set for target user. By simulating on the user-object bipartite network, the initial object resources are reallocated according to trust mechanism. The object's bidirectional diffusion is taken into consideration, and some tunable parameters of object popularity are introduced to achieve resource redistribution, so that the better recommendation results for the target users can be gotten. Experimental results on real-world data sets reveal that the algorithm can enhance the diversity of recommendation results while ensuring higher recommendation accuracy.

[Key words] mass diffusion; trust mechanism; social network; popularity; bipartite network

DOI: 10.19678/j.issn.1000-3428.0049326

0 概述

互联网的迅速发展使得网上信息快速增长, 而导致信息过载。如何快捷有效地从海量的数据中搜寻用户感兴趣的信息是当下面临的一个严峻挑战^[1]。推荐系统成为现在非常流行的信息过滤工具, 为了便于用户在互联网上对海量信息进行筛选, 各式各样的推荐系统应运而生^[2-3]。推荐系统的优势是能够根据用户的行为来捕捉用户的偏好, 从而推荐其感兴趣的物品。近年来, 大量的推荐系统层出不穷, 如目前的推荐系统主要包括基于内容的推荐^[4]、协同过滤推荐^[5]、知识发现推荐^[6]、交互式推荐^[7]以及组合推荐^[8]等。但是, 这些推荐方法在推荐准确率、冷启动和多样性方面依然存在一定的问

题。此外, 基于用户及物品相似度的一些非计算机理论的物理方法也被应用到推荐系统中, 如基于二分网络^[9]的经典物质扩散算法^[10]和热传导算法^[11], 由于其对于用户推荐的准确率高且多样性强而得到广泛应用。

迄今为止, 大部分基于用户-物品二分网络的物质扩散算法都是针对无权网络, 而且对于物品和用户默认都是无差异的^[12], 即在资源扩散的过程中资源都是平均分配到各用户或者物品上, 但是势必会影响推荐的效果, 这是因为不同的用户和物品对于目标用户的影响因子不同。社交网络的崛起已使得互联网上人与人之间交往愈来愈密切, 人与人之间的社交信息在推荐系统中将会起到重要的作用, 通过相似度高的朋友更有利于捕捉目标用户的偏好^[13]。

基金项目: 国家自然科学基金(61371090)。

作者简介: 蔡永嘉(1995—), 男, 硕士研究生, 主研方向为智能信息处理; 李冠宇, 教授、博士、博士生导师。

收稿日期: 2017-11-16 **修回日期:** 2018-02-03 **E-mail:** rabitlee@163.com

可见,用户所处的社交圈非常有利于推荐,而且不同的用户对目标用户影响不一样。物品流行度也会影响目标用户对物品的选择行为。由于传统的物质扩散推荐算法趋于推荐非常流行的物品,而冷门的物品则不易得到推荐,从而导致系统多样性降低。

为解决上述问题,本文基于信任的社交网络,提出一种改进的物质扩散推荐算法(TBIMD)。通过信任机制找到与目标用户相似度高的邻居集,基于用户之间的信任度重新分配物品初始资源,引入参数来调节物品的流行度并对用户资源进行再扩散,进而加强推荐算法的多样性。

1 相关研究

近年来,用户的社交网络信息在互联网推荐系统中扮演着愈发重要的角色,然而在传统的推荐方法中却很少考虑到用户的信任社交网络信息。随着社交网络的飞速发展,基于信任社交网络的推荐将会成为一种十分有效的推荐方式。

融合社交网络信息的协同过滤算法考虑到目标用户的相似度高的好友而不是只考虑与目标用户相似度高的匿名用户集。文献[14]利用与目标用户具有共同爱好的好友信息构建基于目标用户邻居的协同过滤算法,在为目标用户推荐的同时考虑用户的社交信息,但由于社交网络中好友评分很少导致数据稀疏。文献[15]提出基于二分网络的物质扩散推荐算法,该算法假设对象具有一定的初始资源,然后通过二分图进行平均分配扩散最后汇总得到目标用户资源。其通过物质扩散的原理相比传统的协同过滤算法具有更高的精确度,但是未考虑用户的社交信任信息和更倾向推荐热门的物品。文献[16]在用户-物品二分图上呈现一种类似热扩散的热传导算法。该算法将物品看成一个热源,并且同物质扩散类似也在二分图上均匀扩散,最后汇总得到各物品的热量来进行推荐。该算法牺牲了精确度而具有更好的推荐多样性,但是同样未考虑用户的社交网络信息。

综上所述,目前的推荐算法几乎都需要用户对商品的评分、社交网络信息和用户属性等信息对用户进行更准确的推荐。为此,在信任社交网络中,本文提出一种改进的物质扩散算法。该算法考虑了用户的社交网络信任信息,同时改进物质扩散算法来加强推荐的多样性。

2 改进的物质扩散推荐算法

通常推荐系统中经典的物质扩散推荐算法都是基于用户-物品的二分网络。假设推荐系统中包含 m 个用户和 n 个物品,且分别用 $U = \{u_1, u_2, \dots, u_m\}$ 和 $O = \{o_1, o_2, \dots, o_n\}$ 表示用户集和物品集。若用户

u_i 选择物品 o_j ,那么 2 个节点之间就存在一条边 e_{ij} ,因此,推荐系统可以被看作包含 $m + n$ 个节点的无向无权图。

图 1 为经典物质扩散算法在用户-物品二分图中的扩散过程。物质传播主要基于目标用户进行,目标用户已选择过的物品初始值均设置为 1,没有选择过的设置为 0。从初始状态开始,首先将物品的资源扩散到用户,用户的资源值为得到的各物品的总和,例如图中数值 5/6 是物品 1 的初始资源平均扩散给用户 a 和 b 以及物品 3 的初始资源平均扩散给用户 a, b 和 c 的资源权重相加。然后从用户到物品进行扩散,例如数值 5/8 为用户 a 的资源数值平均扩散给物品 1 和 3 以及数值用户 b 的资源平均扩散给物品 1、2、3 和 4 所得,其他数值同理可得。最终由各物品获得的资源数对目标用户进行 top- L 的推荐, L 为给目标用户推荐的资源数量。

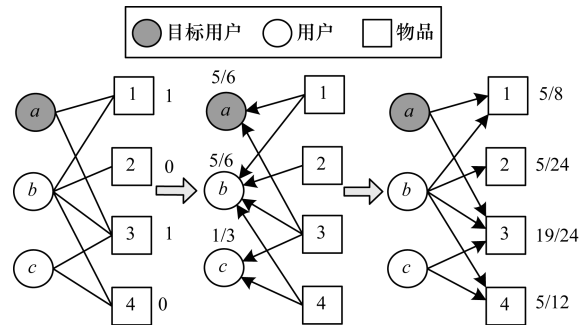


图 1 用户-物品二分图中的物质扩散过程

2.1 算法描述

物质扩散算法的传播规则是将带有一定资源的节点传播到相连的异类节点,而且经典的物质扩散算法是将资源均匀地分配给其他节点。扩散步骤大致分为:物品→用户,用户→物品。

初始化各物品的资源。假设物品 o_β 的初始资源为 f_β ,物品根据流行度 k 平均分给用户 u_i ,用户最终获得所有资源的平均值。计算公式为:

$$r_i = \sum_{\beta=1}^n \frac{e_{i\beta} f_\beta}{k_\beta} \tag{1}$$

其中, k_β 为选择物品 o_β 的用户数,即二分图中物品 o_β 的流行度。然后用户 u_i 根据所得的资源平均分给各物品,最终物品 o_α 获得资源。计算公式为:

$$f'_\alpha = \sum_{i=1}^m \frac{e_{i\alpha} r_i}{k_\alpha} \tag{2}$$

其中, k_α 为选择物品 o_α 的用户数,即二分图中物品 o_α 的流行度。由式(1)和式(2)可得物品 o_β 和物品 o_α 上资源转移权重为:

$$w_{\alpha\beta} = \frac{1}{k_\beta} \sum_{i=1}^m \frac{e_{i\alpha} e_{i\beta}}{k_i} \tag{3}$$

其中, k_β 为选择物品 o_β 的用户数, k_i 为用户 u_i 选择的物品数, $e_{i\alpha}$ 、 $e_{i\beta}$ 分别为用户 u_i 对物品 o_α 、 o_β 的资源

初始值。在经典的物质扩散算法中,两步的扩散过程均选择资源的平均值向对方传播,算法默认用户对各物品的偏好是接近一致的,同时在用户集群中每个物品也是平均扩散资源给相应的用户。实际上,用户对不同的物品喜好程度是有差异的,而且在经典的物质扩散算法中,物品的资源与其流行度成正比,即流行度越高越可能被推荐,从而导致推荐系统的多样性降低。

2.2 改进策略

传统的物质扩散推荐算法没有考虑用户之间以及用户和物品之间的差异。目前兴起的社交网络能够表达用户之间的联系和信息,从而为推荐系统提供更加强有力的手段,相似度高的用户群必定能为目标用户提供更多有用的信息。首先,在基于信任机制的网络上根据信任度找寻与目标用户相似度高的用户。然后,在用户-物品二分图上,为连通的边重新分配初始值,这就使得目标用户没有选择的物品得到了权重,从而有利于提高推荐的多样性。具体表现为:为目标用户推荐自己没有选择而相似度高的邻居群已选择的物品。用户间信任度^[17]计算公式为:

$$T_{a,b} = \frac{d_{\max} - d_{a,b} + 1}{d_{\max}} \quad (4)$$

其中, $d_{a,b}$ 为用户 u_a 和 u_b 在信任网络上的信任传播距离, d_{\max} 为信任网络上允许 2 个用户传播的最大距离,计算公式为:

$$d_{\max} = \frac{\ln n}{\ln k} \quad (5)$$

其中, n 和 k 分别表示在推荐系统中信任网络的范围和平均流行度^[18]。

通过信任度 $T_{a,b}$ 可以确定在用户-物品二分网络中与目标用户相似度高的用户群,利用信任度将二分图中的初始资源进行重新分配。用户 u_i 的初始资源变为:

$$e'_{i\alpha} = e_{i\alpha} + \sum_{k \in U} T_{i,k} \quad (6)$$

其中, $e_{i\alpha}$ 为二分图物品初始值, $T_{i,k}$ 为用户 u_i 和处于信任网络中的其他用户的相似度。将其替换以重新分配物品初始资源,便可更新式(3)。同时,在经典物质扩散算法中对于冷门物品推荐较少,并且许多研究表明经典物质扩散算法仅考虑从收集对象到未收集对象的单向扩散能力,这将会导致不对称估计。为解决上述问题,本文考虑未收集物品向收集物品的反向扩散。反向扩散的计算公式为:

$$w'_{\alpha\beta} = \frac{w_{\beta\alpha}}{\sum_{i=1}^n w_{i\alpha}} \quad (7)$$

本文引入可调参数 λ 、 σ 来抑制过于流行的物品,提高冷门物品的推荐效果。改进的权重转移计算公式为:

$$w'_{\alpha\beta} = \left(\frac{1}{k_{\alpha}^{\lambda}} w_{\alpha\beta} + \frac{1}{k_{\beta}^{\sigma}} w'_{\alpha\beta} \right) / 2 \quad (8)$$

其中, λ 、 σ 为可调参数,用于调节物品的流行度, λ 的取值范围为(0,1), σ 的取值大于0。若不考虑反向扩散且 $\lambda = 0$ 时,则变为传统的物质扩散算法,引入可调参数主要是抑制过于流行的物品,提高冷门物品的推荐率。TBIMD 算法描述如下:

```

输入 datasets fileName
输出 top-L recommendation list
1. calculate trusts; // besides epinions
2. update datasets;
3. readData( fileName );
4. While( uid ∈ Uid )
5. Do{
6. Recommend( uid ); // resource redistribution
7. While( oid ∈ Oid )
8. Do{
9. IMDiffusion( oid ); // resource diffusion
10. }
11. sort( recommend. list );
12. for( int i:L )
13. ofstream ftop-L( top-L );
14. }
15. Get the top-L recommendation list

```

TBIMD 算法首先计算用户之间的信任值,读取数据集后利用 Recommend 函数重新分配物品资源,然后使用 IMDiffusion 函数转移物品资源权重,最后排序获取目标用户的 top-L 推荐列表。该算法通过考虑信任度高的用户以重新分配物品初始资源,再利用双向扩散使得算法灵活可调,可在增强多样性的同时引入可调参数提高算法准确性。

3 实验结果与分析

3.1 实验平台及数据

实验平台运行在 Windows 10 系统上,处理器为 Core™ i5-6200U,内存为 8 GB,实验工具为 Visual Studio 2015。为评估本文提出的 TBIMD 算法的推荐效果和性能,采用 3 个现实生活中的标准数据集,具体见表 1。

表 1 实验数据集来源

数据集	来源网址
MovieLens	https://grouplens.org/datasets/movielens/#attachments
Netflix	http://www.netflixprize.com/
Epinions	http://www.trustlet.org/datasets/downloaded_epinions/

为捕获用户偏好以进行更加准确的推荐,本文将获取数据集中评分 3 星及以上的数据(评分范围为 1 星~5 星),经过处理的数据集信息如表 2 所示。实验前数据集被随机分为 10 份,其中,9/10 为训练集,1/10 为测试集。推荐预测取 10 次结果的平均值作为最终推荐结果。

表 2 实验数据集信息

数据集	用户数	物品数	数据量	稀疏度
MovieLens	943	1 682	82 520	0.630 0
Netflix	10 000	6 000	701 947	0.011 7
Epinions	4 066	7 649	154 122	0.005 0

3.2 度量指标

在推荐系统中,准确性、多样性是评估推荐性能的重要指标。本文引入精确率(P)、平均排序打分(ARS)以及召回率(R),F1 指标(F_1)和平均精度均值(MAP)来评估推荐准确性。同时,引入海明距离(H_{ij})来度量推荐算法的多样性。

1) 精确率

精确率指对于任意一个目标用户生成的 top- L 推荐列表中推荐测试数据集 $T_i(L)$ 所占的比值。计算公式为:

$$P = \frac{1}{m} \sum_{i=1}^m \frac{T_i(L)}{L} \quad (9)$$

2) 平均排序打分

平均排序打分是一种计算推荐命中的物品在推荐列表排名位置的方法^[19]。若用户在测试集 E_p 中搜集了 o_α , 计算其在推荐列表的位置,则在测试集中,所有物品排序分值越小,说明推荐准确度越高。计算公式为:

$$ARS = \frac{1}{|E_p|} \sum_{e_{i\beta} \in E_p} \frac{rank_{i\beta}}{N_i} \quad (10)$$

其中, E_p 为测试集, $rank_{i\beta}$ 为用户 u_i 选择 o_β 所处的排名位置, N_i 为用户未能搜集的物品数。

3) 召回率

召回率是指推荐结果正确命中数和测试集中用户实际选择物品个数的比值^[20], R 值越高推荐效果越好。计算公式为:

$$R = \frac{1}{|E_p|} \sum_{i=1}^m T_i(L) \quad (11)$$

4) F1 指标

F1 指标是用来衡量推荐算法精确度的指标,同时兼顾算法精确率和召回率,是两者的加权平均。计算公式为:

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (12)$$

5) 平均精度均值

平均精度均值表示推荐结果的平均准确率,推荐相关物品越靠前,其值越高。计算公式为:

$$MAP = \frac{\sum_{r=1}^n (P_r \times rel_r)}{N_{rel}} \quad (13)$$

其中, r 表示推荐物品的排名位置, P 为精确率, n 为推荐数量, rel_r 为指定排名 r 的二元函数, N_{rel} 为推荐相关物品的数量。

6) 海明距离

本文采用平均海明距离^[21]度量推荐多样性。

计算公式为:

$$H_{ij} = \frac{1}{m(m-1)} \sum_{i \neq j} \left(1 - \frac{Q_{ij}}{L}\right) \quad (14)$$

其中, Q_{ij} 为用户 u_i 和 u_j 在推荐列表 L 中相同物品的个数。用户之间的海明距离的平均值即表示推荐系统的多样性, H_{ij} 越大,表示推荐的多样性越强。

3.3 结果分析

本文通过实验得到 TBIMD 算法在 MovieLens、Netflix 和 Epinions 这 3 个数据集下最优的 λ 和 σ 值。图 2 表示当推荐列表长度 $L = 50$ 时 TBIMD 算法在不同数据集上精确率和召回率这 2 个度量指标随 λ 和 σ 的变化情况。

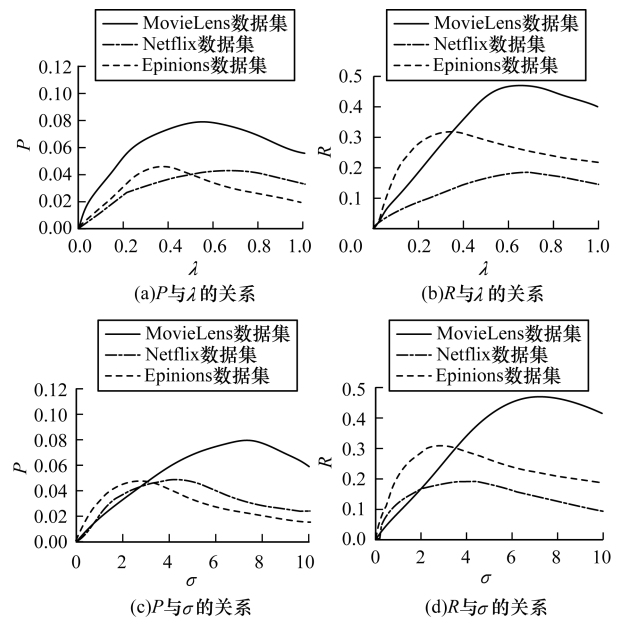


图 2 TBIMD 在 3 个数据集上 P 和 R 随 λ 和 σ 的变化情况

从图 2 可以看出,在 3 个数据集中 λ 分别在 0.62、0.71 和 0.35 附近度量指数达到最优值, σ 分别在 7.75、4.33 和 2.85 附近度量指数达到最优值。在对比实验中,为验证所提算法各度量指标的优越性,将使用最优的可调参数 λ 和 σ 进行实验。由于 σ 远比 λ 要小,表明过于流行的物品对于推荐效果影响较大,因此抑制并且相应增加冷门物品的影响,将会产生更好的推荐结果。

当推荐列表数量 $L = 50$, 参数值(λ, σ)在 3 个数据集中分别取最优值(0.62, 7.75)、(0.71, 4.33)和(0.35, 2.85)时,本文提出的 TBIMD 算法与现有算法的比较结果如表 3 所示。与 TBIMD 算法进行比较的算法包括经典物质扩散算法(MD)、协同过滤算法(CF)^[22]、基于异构网络的推理算法(HNBI)^[23]和基于不平衡一致性的推理算法(UCBI)^[24]。为进一步验证 TBIMD 算法的性能,在参数值最优的情况下,通过改变目标用户推荐列表长度比较各算法性能,实验结果如表 4 所示。

表 3 推荐列表 $L=50$ 时各算法在 3 个数据集上的性能比较结果

性能指标	MoiveLens 数据集					Netflix 数据集					Epinions 数据集				
	MD 算法	CF 算法	HNBI 算法	UCBI 算法	TBIMD 算法	MD 算法	CF 算法	HNBI 算法	UCBI 算法	TBIMD 算法	MD 算法	CF 算法	HNBI 算法	UCBI 算法	TBIMD 算法
P	0.053 7	0.060 9	0.068 4	0.075 1	0.080 6	0.022 4	0.025 5	0.027 5	0.033 9	0.035 2	0.019 7	0.020 1	0.025 4	0.031 9	0.038 1
ARS	0.114 5	0.122 5	0.107 1	0.093 9	0.085 5	0.202 8	0.180 3	0.162 6	0.142 5	0.131 4	0.238 6	0.208 8	0.197 9	0.190 3	0.170 4
R	0.293 3	0.401 1	0.442 4	0.459 3	0.478 5	0.088 1	0.111 2	0.120 7	0.140 3	0.144 9	0.105 1	0.263 1	0.282 4	0.297 7	0.317 8
F_1	0.090 8	0.105 7	0.118 5	0.129 1	0.138 0	0.035 7	0.041 5	0.044 8	0.054 6	0.056 6	0.033 2	0.037 3	0.046 6	0.057 6	0.068 0
MAP	0.052 0	0.059 2	0.069 6	0.073 3	0.082 5	0.020 8	0.025 2	0.028 9	0.036 8	0.037 2	0.020 2	0.024 7	0.026 6	0.032 0	0.038 9
H_{ij}	0.552 7	0.580 1	0.652 8	0.737 9	0.810 6	0.556 7	0.675 5	0.785 9	0.883 1	0.937 7	0.543 3	0.631 6	0.741 9	0.812 5	0.886 9

表 4 推荐列表 $L=20$ 时各算法在 3 个数据集上的性能比较结果

性能指标	MoiveLens 数据集					Netflix 数据集					Epinions 数据集				
	MD 算法	CF 算法	HNBI 算法	UCBI 算法	TBIMD 算法	MD 算法	CF 算法	HNBI 算法	UCBI 算法	TBIMD 算法	MD 算法	CF 算法	HNBI 算法	UCBI 算法	TBIMD 算法
P	0.083 5	0.100 9	0.118 5	0.132 3	0.146 6	0.026 8	0.040 1	0.043 9	0.050 1	0.052 6	0.022 7	0.039 5	0.041 6	0.044 9	0.049 8
ARS	0.114 5	0.122 5	0.107 1	0.093 9	0.085 5	0.202 8	0.180 3	0.162 6	0.142 5	0.131 4	0.238 6	0.208 8	0.197 9	0.190 3	0.170 4
R	0.143 8	0.191 4	0.200 7	0.225 4	0.238 6	0.021 7	0.031 1	0.037 6	0.042 4	0.040 8	0.044 2	0.110 8	0.101 5	0.130 2	0.137 7
F_1	0.1057	0.132 1	0.149 0	0.166 7	0.181 6	0.024 0	0.035 0	0.040 5	0.045 9	0.046 0	0.030 0	0.058 2	0.059 0	0.066 8	0.073 1
MAP	0.080 2	0.097 9	0.117 7	0.129 3	0.141 4	0.020 8	0.043 3	0.045 8	0.049 6	0.055 6	0.022 1	0.040 1	0.042 6	0.045 8	0.050 5
H_{ij}	0.574 4	0.700 8	0.701 9	0.862 2	0.931 3	0.629 1	0.801 4	0.885 9	0.902 2	0.973 1	0.554 3	0.787 6	0.881 6	0.923 3	0.965 9

从表 3 和表 4 可以看出,在不同长度的推荐列表下,对于不同的数据集,TBIMD 算法在推荐准确性和多样性方面均优于现有算法。其中在推荐列表为 50 时,TBIMD 算法较经典物质推荐算法在 3 个数据集上平均排序打分分别提高 25.32%、35.21% 和 28.58%,推荐的多样性分别提高 46.66%、68.44% 和 63.24%。这也表明本文根据用户的信任相似度重新分配物品初始资源和物质双向扩散的方法能够得到更好的推荐结果。

为进一步验证 TBIMD 算法的优越性,本文给出在最优的参数值及推荐物品列表下的精确度-召回率曲线,如图 3 所示,在最优的可调参数值下,TBIMD 算法在不同推荐列表和数据集下的表现均为最优。

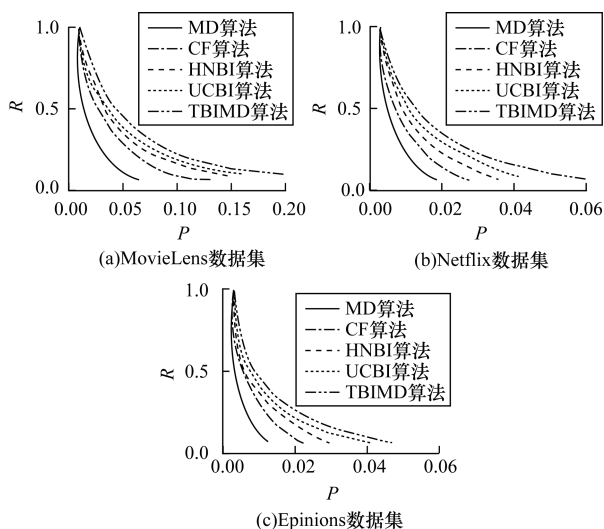


图 3 在 3 个数据集上各算法的精确率-召回率曲线

4 结束语

本文针对现有物质扩散算法存在的问题,提出一种改进的物质扩散算法。由于不同的用户群和物品对目标用户的影响程度不同,因此在信任社交网络中,找出与目标用户信任度高的用户,并且在用户-物品二分图中重新分配物品的初始资源。由于在二分图中单向扩散会导致非对称相似性估计,因此本文引入双向扩散,加入可调参数 λ 和 σ 抑制过于流行的物品,并增加冷门物品的推荐率。实验结果表明,在最优参数下,本文提出的 TBIMD 算法提高了推荐系统的准确率和多样性。目前,基于信任的社交网络虽然可以搜寻接近目标用户的相似用户群,但还有一些重要因素尚未考虑,如用户情景因素等,因此下一步将对一些未考虑但对推荐系统具有重要影响的因素进行分析研究。此外,如何在社交网络中挖掘用户的隐式个性化信息,并将其引入到推荐系统中,以获取更好的推荐效果,也是下一步研究的重要方向。

参考文献

- [1] ZHOU T, KUSCSIK Z, LIU J G, et al. Solving the apparent diversity-accuracy dilemma of recommender systems[J]. PNAS, 2010, 107(10): 4511-4515.
- [2] GOSWAMI S. Analysing effects of information overload on decision quality in an online environment[J]. Journal of Management Research, 2015, 15(4): 231-245.
- [3] HE C, PARRA D, VERBERT K. Interactive recommender systems: a survey of the state of the art and future research challenges and opportunities [J]. Expert Systems with Applications, 2016, 56: 9-27.

- [4] PESSEMIER T D, COPPENS S, GEEBELEN K, et al. Collaborative recommendations with content-based filters for cultural activities via a scalable event distribution platform[J]. *Multimedia Tools and Applications*, 2012, 58(1):167-213.
- [5] 黄创光, 印鉴, 汪静, 等. 不确定近邻的协同过滤推荐算法[J]. *计算机学报*, 2010, 33(8):1369-1377.
- [6] SIPPEL S. Knowledge-based recommendations for cocktail recipes[EB/OL]. [2017-10-14]. <http://www.10.1056/NEJM199809173391208>.
- [7] GARG N, WEBER I. Personalized, interactive tag recommendation for flickr [C]//*Proceedings of ACM Conference on Recommender Systems*. New York, USA: ACM Press, 2008:67-74.
- [8] BRAUNHOFER M, CODINA V, RICCI F. Switching hybrid for cold-starting context-aware recommender systems [EB/OL]. [2017-10-14]. <http://www.10.1145/2645710.2645757>.
- [9] 周波, 杨朝峰. 基于二分网络的推荐算法研究进展[J]. *情报工程*, 2016, 2(1):77-90.
- [10] ZHOU T, LÜ L, ZHANG Y C. Predicting missing links via local information [J]. *European Physical Journal B*, 2009, 71(4):623-630.
- [11] GUO Q, SONG W J, HOU L, et al. Effect of the time window on the heat-conduction information filtering model [J]. *Physica A: Statistical Mechanics and Its Applications*, 2014, 401(5):15-21.
- [12] LIU J G, HU Z, GUO Q. Effect of the social influence on topological properties of user-object bipartite networks [J]. *European Physical Journal B*, 2013, 86(11):1-11.
- [13] XU G, WU Z, ZHANG Y, et al. Social networking meets recommender systems[EB/OL]. [2017-10-14]. <http://www.10.1504/IJNSM.2015.069773>.
- [14] CAI X, BAIN M, KRZYWICKI A, et al. Collaborative filtering for people to people recommendation in social networks[J]. *International Journal of Human-Computer Studies*, 2015, 76(C):50-66.
- [15] ZHOU T, REN J, MEDO M, et al. Bipartite network projection and personal recommendation [J]. *Physical Review E: Statistical, Nonlinear and Soft Matter Physics*, 2007, 76(2).
- [16] LIU J G, ZHOU T, GUO Q. Information filtering via biased heat conduction [J]. *Physical Review E: Statistical, Nonlinear and Soft Matter Physics*, 2011, 84(3).
- [17] MASSA P, AVESANI P. Trust-aware recommender systems [C]//*Proceedings of ACM Conference on Recommender Systems*. New York, USA: ACM Press, 2007:17-24.
- [18] YUAN W, GUAN D, LEE Y K, et al. Improved trust-aware recommender system using small-worldness of trust networks [J]. *Knowledge-Based Systems*, 2010, 23(3):232-238.
- [19] LÜ L, LIU W. Information filtering via preferential diffusion [J]. *Physical Review E: Statistical, Nonlinear and Soft Matter Physics*, 2011, 83(2).
- [20] DAVIS J, GOADRICH M. The relationship between precision-recall and ROC curves [C]//*Proceedings of International Conference on Machine Learning*. New York, USA: ACM Press, 2006:233-240.
- [21] ZHU X, TIAN H, CAI S. Personalized recommendation with corrected similarity [EB/OL]. [2017-10-14]. <https://arxiv.org/abs/1405.4095>.
- [22] YANG X, GUO Y, LIU Y, et al. A survey of collaborative filtering based social recommender systems [J]. *Computer Communications*, 2014, 41(5):1-10.
- [23] ZHOU T, JIANG L L, SU R Q, et al. Effect of initial configuration on network-based recommendation [J]. *Physics*, 2008, 81(5).
- [24] ZHU X, TIAN H, ZHANG P, et al. Personalized recommendation based on unbiased consistence [J]. *Europhysics Letters*, 2015, 111(4).

编辑 陆燕菲

(上接第181页)

- [10] KIM S, YANG S, KO Y. Mathematical equation retrieval using plain words as a query [C]//*Proceedings of the 21st ACM International Conference on Information and Knowledge Management*. New York, USA: ACM Press, 2012:2407-2410.
- [11] LIŠA M, SOJKA P. Combining text and formula queries in math information retrieval: evaluation of query results merging strategies [C]//*Proceedings of the 1st International Workshop on Novel Web Search Interfaces and Systems*. New York, USA: ACM Press, 2015:7-9.
- [12] STATHOPOULOS Y, TEUFEL S. Mathematical information retrieval based on type embeddings and query expansion [EB/OL]. [2018-08-06]. <https://www.aclweb.org/anthology/C/C16/C16-I221.pdf>.
- [13] PATHSK A, PAKRAY P, SARKAR S, et al. MathIRs: retrieval system for scientific documents [J]. *Computación Y Sistemas*, 2017, 21(2):253-265.
- [14] 贺海涛, 郑山红, 李万龙, 等. 基于关联规则和语义规则的本体概念提取研究[J]. *吉林大学学报(信息科学版)*, 2014, 32(6):657-663.
- [15] 刘剑, 唐慧丰, 刘伍颖. 一种基于统计技术的中文术语抽取方法[J]. *中国科技术语*, 2014(5):10-14.
- [16] 刘群, 张华平. 基于层叠隐马模型的汉语词法分析[J]. *计算机研究与发展*, 2004, 41(8):1421-1429.
- [17] 周浪, 张亮, 冯冲, 等. 基于词频分布变化统计的术语抽取方法[J]. *计算机科学*, 2009, 36(5):177-180.
- [18] 董洋溢, 李伟华, 于会. 文本特征和复合统计量的领域术语抽取方法[J]. *西北工业大学学报*, 2017, 35(4):729-735.
- [19] 任禾, 曾隽芳. 一种基于信息熵的中文高频词抽取算法[J]. *中文信息学报*, 2006, 20(5):40-43, 90.
- [20] 李丽双, 王意文, 黄德根. 基于信息熵和词频分布变化的术语抽取研究[J]. *中文信息学报*, 2015, 29(1):82-87.
- [21] 何思谦, 刘绍学, 朱元森, 等. *数学辞海* [M]. 太原: 山西教育出版社, 2002.
- [22] 秦玉平, 唐亚伟, 伦淑娟, 等. 一种基于二叉树的数学公式匹配算法[J]. *计算机科学*, 2013, 40(5):251-252, 278.
- [23] LaTeX search [EB/OL]. [2018-08-06]. <http://www.latexsearch.com/>.

编辑 赵辉