

doi:10.3772/j.issn.2095-915x.2016.02.009

发送者和接受者能力的 二分网络推荐算法研究

周波, 杨朝峰

(中国科学技术信息研究所 北京 100038)

摘要: 在大数据时代, 推荐系统在电子商务等领域中发挥着越来越重要的作用。本文通过对现有的基于二分网络推荐算法进行分析总结, 提出了 TAA 算法, 并将该算法同物质扩散算法、热传导算法、偏热传导算法、混合算法和接收者能力算法在 Movielens 数据集上进行对比实验, 实验结果表明该算法推荐结果的准确率和召回率均优于现有的 5 种算法。

关键词: 二分网络, 推荐算法, TAA 算法

The research of Bipartite Network Recommendation Algorithm Based on Transmitter and Acceptor Ability

ZHOU Bo, YANG ChaoFeng

(Institute of Scientific and Technical Information of China, Beijing 100038, China)

Abstract: In the time of big data, recommendation system plays an important role in e-commerce and other fields. This paper analyzed the existed recommendation algorithms based on bipartite network, and proposed a new algorithm based on transmitter and acceptor ability (TAA) algorithm. Compared with the material diffusion algorithm, heat conduction algorithm, partial Heat conduction algorithm, combination algorithm and receiver Capacity algorithm, the performances of TAA algorithm indicated that it can get a higher precision rate and recall rate.

Keywords: Bipartite network, recommendation algorithm, TAA algorithm

作者简介: 周波 (1991-), 中国科学技术信息研究所情报学硕士研究生, 研究方向: 数据挖掘与知识组织, E-mail: hubeizhoubo@163.com; 杨朝峰 (1975-), 中国科学技术信息研究所副研究员, 同济大学技术经济及管理专业博士, 研究方向: 科技政策、宏观经济政策。

1 引言

推荐系统在信息大爆炸时代的重要性不言而喻。在电商领域里,推荐系统可以直接增加销售收入,亚马逊据称有20%~30%的销售收入来自于推荐系统^[1],阿里更是每年举办推荐算法大赛来提高推荐准确率。在大数据时代,信息的供需双方更加需要推荐系统,对于信息的供应方来说,面对众多消费者、千变万化的市场机遇和碎片化的消费需求,如何将最有价值的信息推送给最需要的人,这是供方的痛点;对需求方来说,如何从海量的搜索结果中找到合适的信息,是需求方的痛点。

为解决大数据时代信息供需双方的痛点,大量的研究者对推荐系统展开了研究,产生了大量有效的推荐算法。这些研究主要可以分为以下三个方面,基于内容的推荐算法、基于结构的推荐算法和推荐算法融合。基于内容的推荐技术的核心思想是分析用户历史内容,根据用户和对象内容的相似度来进行推荐,基本思路是先进行文本挖掘,然后根据文本的相似性来进行推荐,具体的有基于TF-IDF的文本分析推荐技术^[2],基于主题模型的文本推荐技术^[3],基于个人知识地图的推荐技术^[4],基于海量内容的推荐技术等^[5]。基于结构的推荐技术有:协同过滤推荐技术^[6]、基于网络结构的推荐技术^[7]和基于关联规则的推荐技术等。融合推荐的基本思想是通过融合来弥补各类算法的缺点,发挥各类算法的优点,达到提高推荐准确率的效果。

复杂网络是新兴的交叉学科,涵盖了数学、物理、计算机,人工智能等领域的知识,同时又对这些领域的研究起到推动作用。复杂网络产生的理论和方法对数据挖掘,机器学习以及推荐算法等产生了较大的影响^[8]。基于二分网络的推荐算法由周涛^[9]于2007年基于复杂网络理论而提出,该算法提出后,受到广泛关注,基于二分网

络的推荐算法较传统的协同推荐算法有着更高的推荐精度,此后衍生出一系列的基于二分网络的推荐算法以及各种优化算法。这些优化的推荐算法以物质扩散算法和热传导算法为基础,以后的研究主要针对具体的情况做出优化,并且都取得了不错的效果。本文对以往的几种具有代表性的二分网络推荐算法进行分析研究,在以往的算法基础上提出了发送者和接受者能力的推荐算法,并进行了实验对比验证。

2 二分网络推荐算法概述

目前,基于二分网络的推荐算法具有代表性的有以下五种:

第一种是物质扩散算法,该算法假设每个对象均有一定的初始资源,通过对象的度将资源平均地分配给相邻的用户,然后每个用户又将自己所有分到的资源再次平均地分配给所选择的对象,通过汇总对象的所有相邻用户分配的资源,得到该对象获得的资源。现已经证明通过初始资源在网络上的扩散原理进行链路预测,比协同过滤算法具有更高的精度^[1]。基于物质扩散的算法,由于不考虑用户和对象的内容特征和思想特征,只是把他们抽象成二分网络的节点,算法仅利用隐藏在用户和物品的选择关系之中的信息。文献^[10]中,在二分网络中引入了扩散动力学,实现了基于物质扩散(MD)算法,证明了这些算法的结果明显好于经典的协同过滤。

用 X, Y 分别表示二分网络两类节点用户和物品的集合,其中 X 类节点有 m 个, Y 类节点有 n 个, A 表示 X 同 Y 的连接关系, B 是 A 的转置矩阵,如果 X_i 同 Y_j 相连,那么 $a(i,j)=1, b(j,i)=1$,否则为零, $d(X_i)$ 和 $d(Y_j)$ 分别表示的是用户 X_i 和物品 Y_j 的度。(下文相关的公式符号含义同此)设用户 X_i 对物品 Y_j 的初始资源为 $h(j,i)$,物品将资源平均的分配给每个用户,每个用户又将所得资源平均分配

给他选择过的物品。

矩阵 $W(kj)$ 记为资源分配矩阵, 表示物品 j 愿意分配给物品 k 的资源配额, 则物质扩散算法的资源分配矩阵 $W(kj)$ 的一般表达式如下, 使用 $W(kj) * h(ji)$ 得到用户 i 对物品 K 的资源值。用户 i 对所有物品的资源值进行排序, 取前 $top-N$ 作为推荐结果。

$$W(kj) = \frac{1}{d(Yj)} \sum_{l=1}^m \frac{b(k,l) * b(j,l)}{d(Xl)}$$

第二种是热传导算法, 基于热传导的链路预测方法其思想来自于物理学的热传导链路预测过程, 但是其将物理学中的热传导方程进行离散化, 物理学中的热平衡方程是 $-k \nabla^2 T(r) = \nabla * J(r)$, k 代表导热性, $\nabla^2 T(r)$ 代表温度梯度^[11]。用物理学中的基本理论来进行链路预测存在两个问题, 一是算法的时间复杂度和空间复杂度都很高, 影响大规模的数据处理和运算, 二是理解起来较为复杂。鉴于此, 文献 [12] 根据热传导的思想提出了一种更加形象的热传导算法, 基本形式同热传导算法相似, 但是有区别, 热传导算法资源分配矩阵 $W(kj)$ 的一般表达式为:

$$W(kj) = \frac{1}{d(Yk)} \sum_{l=1}^m \frac{b(k,l) * b(j,l)}{d(Xl)}$$

第三种是偏热传导算法。热传导算法会降低冷门的物品推荐的力度, 强化推荐系统的马太效应, 如何降低马太效应同时提高推荐精度就是偏热传导算法的主要出发点。文献 [13] 考虑了这种情况, 并以热传导算法为基础, 提出了偏热传导算法, 基本原理同热传导算法相似, 但是考虑大度物品的影响力, 在最后一步传导过程中考虑物品度的影响, 物品最后得到资源不是除以物品的度, 而是除以物品度的 λ 次方, λ 是一个可调节的参数, 这样热传导算法的资源分配矩阵为:

$$W(kj) = \frac{1}{d(Yk)^\lambda} \sum_{l=1}^m \frac{b(k,l) * b(j,l)}{d(Xl)}$$

第四种是接受者能力算法。文献 [14] 研究了资源接受者接收能力的差异对推荐精度的影响。例如, 将一滴墨水滴入水杯中同将一滴水滴入大海中, 其影响效果就截然不同, 将一滴水滴入水杯中, 水杯中水的颜色就改变了, 而滴入大海中却没有任何变化。考虑这种情况对推荐算法的影响, 基于资源接收能力不同的个性化算法被提出。资源传播矩阵 $W(kj)$ 表示物品 Yk 从物品 Yj 那里得到的资源数量, 在资源扩散过程中, 考虑接收者 Yk 对所收到资源的响应, 假定响应反比于它们自身度的 λ 次方, 则得到资源分配矩阵为:

$$W(kj) = \frac{1}{d(Yk)^\lambda d(Yj)} \sum_{l=1}^m \frac{b(k,l) * b(j,l)}{d(Xl)}$$

第五种是基于融合推荐算法, 物质扩散和热传导结合的混合推荐算法, 物质扩散算法和热传导算法各具特点, 物质扩散算法推荐的准确性更高, 适合用户找到自己感兴趣的物品, 而热传导算法具有较强的冷门商品推荐能力, 推荐结果的多样性较好, 更容易为用户提供个性化的物品。因此可以考虑将两者进行融合以提高推荐效果和推荐质量。因此, 文献 [13] 物质扩散算法和热传导算法的资源转移方程的相似性提出了将二者融合的算法, 资源转移方程如下:

$$W(kj) = \frac{1}{d(Yk)^\lambda d(Yj)^{1-\lambda}} \sum_{l=1}^m \frac{b(k,l)b(j,l)*}{d(Xl)} \quad (0 \leq \lambda \leq 1)$$

该算法与热传导和物质扩散算法只是资源转移方程不同, 其他的相同。

基于二分网络的推荐算法的实质, 以某个节点为初始节点, 然后使用一定的资源分配规则, 向相邻的节点分配资源, 分配的资源越多的节点,

被认为是同初始节点的关系越密切,被推荐的力度应该越大。基于此,各个推荐算法的不同之处就在于资源分配规则的不同。

下面就对以往的各种推荐算法的资源分配规则进行分析。物质扩散算法的资源分配规则是将自身按照其自身的度向各个节点进行平均分配,节点自身的度越小其资源分配能力就会越强,热传导的资源分配规则是按照资源分配对象的度进行平均分配,分配对象的度越小其资源分配能力就会越强。偏热传导算法的资源分配规则是建立在物质扩散算法的基础上的,第一个节点是按照自身度向相邻的节点进行平均分配,而第二个节点却是按照自身度的幂次方向相邻的节点进行资源分配。基于接受者能力不同的推荐算法的资源规则是在物质扩散算法的基础上,在最后一步资源分配中,接受者接收到的资源先是发送者根据自身的度进行均分,然后再按照接收者度的幂次方均分。基于融合的推荐算法的资源分配规则是将物质扩散算法和热传导算法相结合,节点所接受到的资源不仅同自身的度数相关而且同发送者的度数相关。

如果从接受者和发送者的角度来看,物质扩散算法只是考虑了资源发送者能力的推荐算法,热传导算法只是考虑了资源接受者能力的推荐算法,偏热传导算法是对资源发送者的能力来设计推荐算法,基于融合的推荐算法是将发送者和接受者能力综合起来考虑的推荐算法,基于接受者接收能力不同的推荐算法也是发送者和接受者能力不同的推荐算法,但其只是考虑了在最后一步中接受者能力的推荐算法。

然而,目前需要同时考虑接受者和发送者能力的推荐算法存在融合不彻底的情况。经上文分析,同时考虑接受者和发送者能力的推荐算法有用户接受者能力的推荐方法和基于融合的推荐算法,基于用户接受者能力的推荐方法只是在最后

一步的资源扩散中考虑将发送者能力和接受者能力结合起来确定最终的资源分配;基于融合的推荐算法严格意义上讲其并不将接受者和发送者能力进行融合,而是两种算法的融合,而且其存在这样一种情况,发送者资源扩散能力强后接受者接收的能力就弱,资源接受者接收的能力强的话,资源分配者的能力就弱,基于融合的推荐算法的推荐算法相当于是在做这样一个假设,发送者能力和接受者能力相互关联。笔者认为这种情况并不符合实际情况,资源发送者和资源接受者能力并不存在此强彼弱的关系,实际上推荐系统所面对的用户是多种多样的,需求也是多种多样的,基于融合的推荐算法没有考虑到接受者能力和接受者能力都很弱及接受者能力和接受者能力都很强的情况。

因此,本文对以往的二分网络推荐算法进行总结归纳提出更为彻底的接受者能力和发送者能力的推荐算法,称为TAA(Transmitter and Acceptor Ability Algorithm)算法,简单来说就是在每一次资源扩散中,资源接受者能够接收的资源量是由发送者的能力、接受者的能力所共同决定的,且每次资源扩散的参数是不同的。

3 TAA 算法设计与实验

3.1 TAA 算法设计

使用TAA算法对物品进行推荐的两个步骤,可表示为:

第一步:资源由物品 Y_j 向用户 X_l 扩散,设用户 X_l 的分配的资源为 $f(X_l)$, 则

$$f(X_l) = \sum_{j=1}^n \frac{b(j,l) * h(j)}{d(Y_j)^{\lambda_1} d(X_l)^{\lambda_2}}$$

第二步:资源由用户 X_l 向物品 Y_k 扩散,物

品 Yk 所得的资源为 g(Yk),

$$g(Yk) = \sum_{l=1}^m \frac{f(Xl) * b(k, l)}{d(Xl)^{\lambda 3} d(Yk)^{\lambda 4}}$$

矩阵 W(kj) 记为资源分配矩阵, 表示物品 j 愿意分配给物品 k 的资源配额, 则 W(kj) 的一般表达式为:

$$\begin{aligned} W(kj) &= \frac{1}{d(Xl)^{\lambda 3} d(Yk)^{\lambda 4}} \sum_{l=1}^m \frac{b(k, l) * b(j, l)}{d(Yj)^{\lambda 1} d(Xl)^{\lambda 2}} \\ &= \frac{1}{d(Yj)^{\lambda 1} d(Yk)^{\lambda 4}} \sum_{l=1}^m \frac{b(k, l) * b(j, l)}{d(Xl)^{\lambda 2 + \lambda 3}} \\ &= \frac{1}{d(Yj)^{\lambda 1} d(Yk)^{\lambda 3}} \sum_{l=1}^m \frac{b(k, l) * b(j, l)}{d(Xl)^{\lambda 2}} \end{aligned}$$

其中, $-1 \leq \lambda 1 \leq 1; -2 \leq \lambda 2 \leq 2; -1 \leq \lambda 3 \leq 1$

从上式可以看出, 物质扩散算法、热传导算法、基于流行性的算法、基于用户接收能力的算法、基于融合的算法、基于用户所选择物品特性的算法都是基于节点度的 TAA 算法的特例, 表 1 展示了上述各种算法在 TAA 算法下, 各个参数的取值情况。

表 1 TAA 算法参数取值范围

发送者和接收者能力算法	$-1 \leq \lambda 1 \leq 1; -2 \leq \lambda 2 \leq 2; -1 \leq \lambda 3 \leq 1$
物质扩散算法	$\lambda 1=1, \lambda 2=1, \lambda 3=0$
热传导算法	$\lambda 1=0, \lambda 2=1, \lambda 3=1$
偏热传导算法	$\lambda 1=0, \lambda 2=1, -1 \leq \lambda 3 \leq 1$
基于接受者接收能力的算法	$\lambda 1=1, \lambda 2=1, 0 \leq \lambda 3 \leq 1$
基于融合的算法	$\lambda 1 + \lambda 3=1; \lambda 2=1$

3.2 六种推荐算法对比实验

为了验证 TAA 算法的有效性, 将 TAA 算法物质扩散算法、热传导算法、偏热传导算法、接受者接受能力算法和混合算法进行对比实验。

实验采用推荐算法中经典的 movielensMovielens

公开数据集, MovieLens 数据集中, 用户对自己看过的电影进行评分, 分值为 1 ~ 5。MovieLens 包括两个不同大小的库, 适用于不同规模的算法。

小规模库是 943 个独立用户对 1682 部电影作的 10000 次评分的数据, 大规模库是 6040 个独立用户对 3900 部电影作的大约 100 万次评分, 本文采用小规模的数据集。同时本文不预测评分, 因此将评分数据给除掉。生成 5 组随机数据集, 按 80% 和 20% 生成测试集合训练集。

算法伪代码如下:

第一步: 读入数据, 生成用户—电影矩阵 A, 用户的度和电影度的向量 d(X) 和 d(Y), 初始资源分配矩阵 H。

第二步: 生成电影—电影资源扩散矩阵 W。

第三步: 生成预测结果矩阵 D=W*H。

第四步: 在矩阵 D 各列中按资源值的多少进行排序, 截取前 50, 作为推荐结果。

第五步: 对实验结果进行评价。

实验环境为: windows server 2008 R2 Enterprise 的操作系统, intel xeon 2.00GHz 的 CPU, 32.0GB 内存, 64 位操作系统, 编程语言是 R。

推荐算法结果的评价指标最常用的指标有两个: 准确率和召回率, 本文将使用这两个指标来进行推荐结果的评价。准确率反映的是推荐结果中有多大的比例是正确的, 召回率反映的推荐结果中有多大的比例在测试集中出现过。表 2 定义了推荐结果的准确率和召回率。

表 2 推荐算法评价指标

	测试集中	非测试集中	
推荐结果	a	b	推荐结果的 数量 =a+b
推荐结果之外	c	d	
	测试集的 数量 =a+c		

准确率的定义为: $\text{precision} = a/a+b$

召回率的定义为: $\text{recall} = a/a+c$

4 实验结果及讨论

4.1 实验结果

物质扩散算法和热传导算法不用确定最优参数, 而偏热传导算法、接收者接受能力算法、混合算法和 TAA 算法的推荐结果是随着参数变化而变化, 因此需先确定最优参数。

偏热传导算法的可变参数是 λ_3 , 实验结果的结果如下, 可以发现准确率和召回都是随着 λ_3 的增大而先增大后减小, 当 λ_3 等于 0.2 时推荐结果最佳。

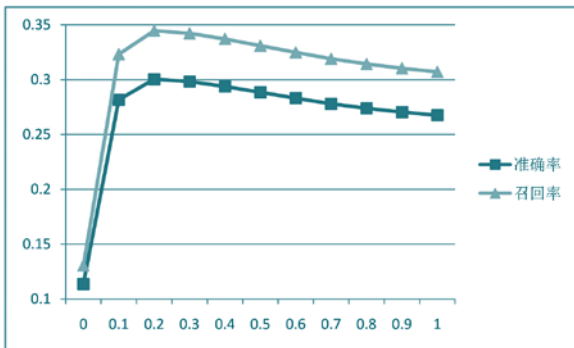


图1 偏热传导算法的实验结果

接收者能力算法的可变参数也是 λ_3 , 实验结果的结果如下, 可以发现准确率和召回都是随着 λ_3 的增大而先增大后减小, λ_3 等于 0.7 时推荐结果最佳。

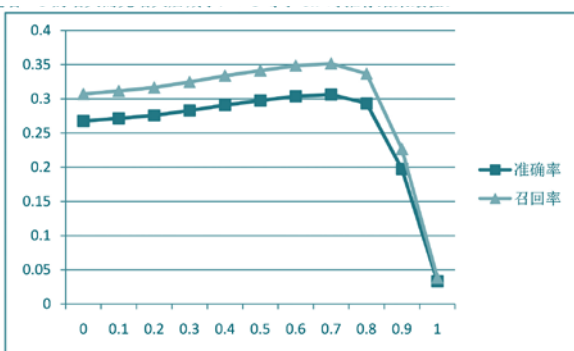


图2 接收者能力 TAA 算法的实验结果

混合算法的可变参数是 λ_1 、 λ_3 , $\lambda_1 + \lambda_3 = 1$ 其中实验结果的结果如下, 可以发现准确率和召回都是随着 λ_1 和 λ_3 的增大而先增大后减小, $\lambda_1 = 0.2$, $\lambda_2 = 1$, $\lambda_3 = 0.8$ 时推荐结果最佳。

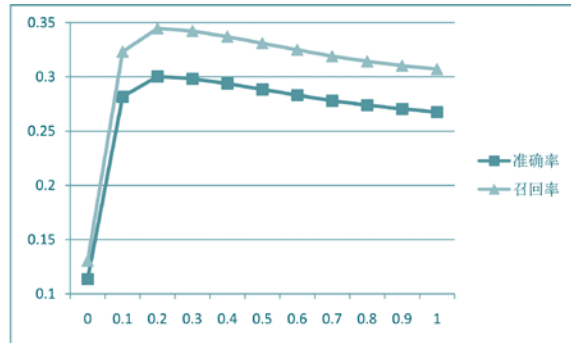


图3 混合算法的实验结果

TAA 算法由于有三个参数, 实验结果数量较多, 为方便展示, 选择实验结果较好的两组进行展示。通过对结果分析可以发现当 λ_3 等于 0.7 和 0.8 时推荐结果较好, 当 λ_3 等于 0.7 时推荐结果分别如图 4 和图 5 (为方便 3 维作图, 已 x 、 y 、 z 轴扩大 100 倍, x 轴代表 λ_1 , y 轴代表 λ_2 , z 轴代表推荐结果)。在 $\lambda_1 = 0.9$, $\lambda_2 = 0.9$ 时推荐结果达到最大, 距离 $\lambda_1 = 0.9$, $\lambda_2 = 0.9$ 越远, 推荐结果的准确率和召回率都是较低。

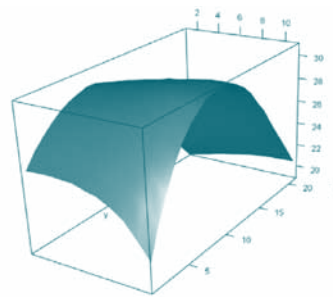


图4 3为0.7时TAA算法的准确率

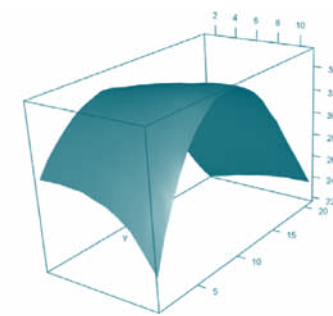


图5 3为0.7时TAA算法的召回率

当 λ_3 等于 0.8 时推荐结果分别如下, 对结果进行分析后发现, 当 $\lambda_1=0.5$, $\lambda_2=0.9$ 时推荐结果最好, 距离该参数越远, 推荐结果越差。 λ_3 等于 0.8 推荐结果的波动性比 λ_3 等于 0.7 的波动要大, 图形的平滑性也较差。

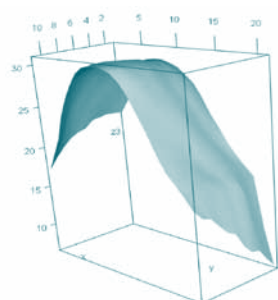


图 6 3 为 0.8 时 TAA 算法的准确率

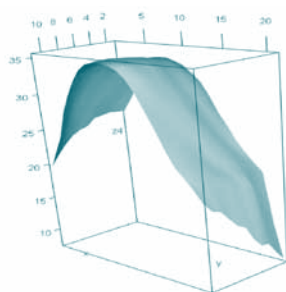


图 7 3 为 0.8 时 TAA 算法的召回率

TAA 算法同其他算法的结果对比 (热传导算法由于准确率和召回率较低未列出), TAA 算法准确率和召回率大幅优于热传导算法、混合算法、偏热传导算法和物质扩散算法, 准确率和召回率分别高于接受者能力算法 0.2% 和 0.225%。

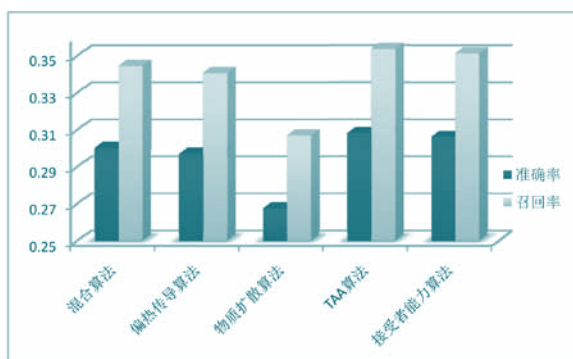


图 8 TAA 算法同其他 4 种算法推荐效果对比

如下表 3 是推荐结果的准确率和召回率的排名, TAA 算法有 13 组参数的推荐结果排名前 13, 接受者能力算法排名 14, 这表明 TAA 算法优于接受者能力算法不是偶然的。

表 3 6 种推荐算法实验结果及排名

算法名称	λ_1	λ_2	λ_3	准确率	召回率	排名
TAA 算法	0.9	0.9	0.7	0.30828	0.35375	1
接受者能力算法	1	1	0.7	0.30632	0.3515	14
混合算法	0.2	1	0.8	0.300349	0.34465	87
偏热传导算法	0	1	0.8	0.297037	0.34085	149
物质扩散算法	1	1	0	0.267582	0.30705	873
热传导算法	0	1	1	0.113725	0.1305	2262

实验结果表明 TAA 算法均优于其他 5 种算法, TAA 算法优于其他算法比率如下表 4。

表 4 TAA 算法优于其他算法比率

算法名称	TAA 算法优于其他算法比率
热传导算法	171.07%
物质扩散算法	15.21%
偏热传导算法	3.78%
混合算法	2.64%
接受者能力算法	0.64%

4.2 实验讨论

通过采用 movielens 数据集, 将 TAA 算法同热传导、物质扩散算法、偏热传导算法、混合算法和接受者能力算法进行了试验, 实验结果表明, 本文提出的 TAA 算法的准确率和召回率均优于其他 5 种算法。从最终结果的排名中可以看到推荐热传导算法与物质扩散算法以及热传导算法的准确率和召回率是最差的, 在第三节中分析中指出, 物质扩散算法是只考虑资源发送者能力的算法,

而热传导算法是只考虑了资源接收者能力的算法，通过对比可以发现这种单一的考虑资源发送者能力或资源接受者能力的算法推荐结果较差。当然，这里热传导算法的推荐结果比物质扩散算法的推荐结果差，并不是说明考虑资源发送者能力的算法比考虑资源接收者能力的算法好，文献 [15] 指出，热传导算法和物质扩散算法的适用范围是不同的，物质扩散算法有助于推荐流行性物品，而热传导算法有助于推荐长尾物品。偏热传导在热传导算法的基础做了一些改进，只考虑了资源接受者能力的算法，只是最后一步资源传递时认为，节点的资源传递能力同节点的度呈指数关系，推荐结果有一提升。

混合算法和接收者能力算法的推荐结果的准确率和召回率较高，较高的原因是在于这两种算法都在一定的限制条件下同时考虑了接受者能力

和发送者能力。如在第三节中指出，混合算法虽同是考虑节点的资源接受能力和发送能力，但是其假设接受者能力和发送者能力存在此强彼弱的关系 ($\lambda_1 + \lambda_3 = 1$ 即是这种假设的证明)，这制约了其推荐结果的提升，实验证明了这种假设不合理。接收者能力算法在物质扩散的基础上，在最后一步的资源扩散中考虑了节点的资源接受能力，由于物质扩散算法是只考虑节点的资源发送能力，因此接受者能力算法相当是在最后一步的资源扩散中同时考虑节点的资源接受能力和发送能力，而第一步中只考虑节点的资源发送能力而没有考虑节点的资源接受能力，因此，其推荐效果不如 TAA 算法，又由于其结构最接近 TAA 算法，因此其推荐结果比 TAA 算法略低。

通过上述实验和分析，可以得出推荐效果同节点的接受能力和发送能力的关系，如下图所示。

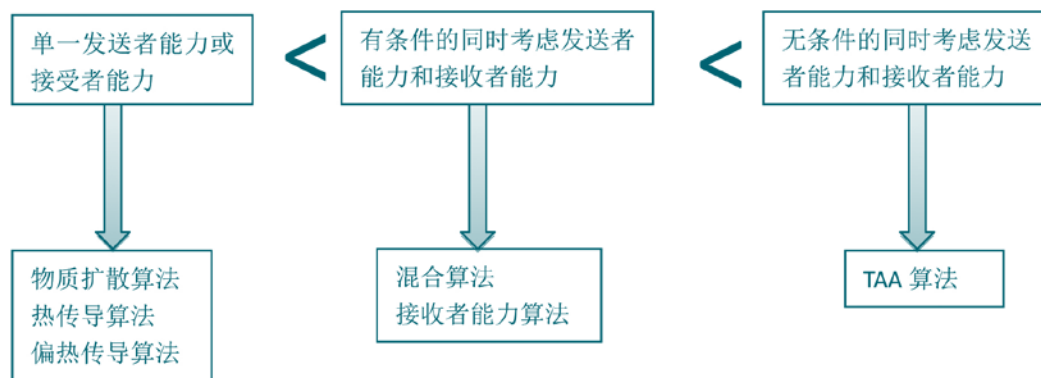


图9 推荐效果同接收者能力和发送者能力

TAA 算法的最优参数 $\lambda_1=0.9$, $\lambda_2=0.9$, $\lambda_3=0.7$ 适当降低大度节点的资源扩散能力和接收能力有助于提高推荐结果的精确率，实验结果同以往的推荐研究结论一样^[16]。与以往不同的是，以往只是在最后一步中降低大度节点的推荐力度来证明这个结论的，而本文的实验证明在每一步中，降低大度节点的推荐力度都有助于提高推荐结果。

5 总结

本文成果如下：

1. 提出来在每一步中同时考虑节点的发送能力和接受能力的思想，并基于此设计了基于发送者和接受者能力的二分网络推荐算法(TAA算法)。
2. 设计实验将 TAA 算法同其他 5 种算法进行了对比，实验结果证明 TAA 算法要优于其他的

5种算法,从而证明了本文所提的TAA算法推荐效果较好。

3. 分析TAA算法推荐效果比其他算法的较好的原因:其他算法是单一或者有限制的同时考虑节点的发送能力和接受能力的算法,而TAA算法是无条件同时考虑节点的接受能力和发送能力的算法。

本文的不足:

1. 未能使用更多的公开数据集进行实验。由于实验设备和计算机计算能力有限,以及TAA算法有三个独立参数需进行大量实验,因此本文仅使用了movielens数据集进行了试验,而没有使用Netflix等数据集进行实验,实验结论缺乏大量的数据集进行验证,这是本文的遗憾。

2. 由于TAA算法有三个独立参数,如果每个参数变化的间隔是0.1则需进行2000次实验,而如果是0.01的话则需进行200万次实验,由于计算机性能因素,本文参数变化的间隔是0.1,这导致本文最优参数只能精确到0.1。如果采用更细的参数间隔的话,本文的最优参数可能会改变,推荐结果可能更好。

未来的研究可在本文的基础上展开如下几方面的研究:

1. 使用更多的数据集,更加精确的参数来验证TAA算法。

2. 使用矩阵分解等技术提升TAA算法性能。大数据环境下,推荐算法的处理速度犹未重要尤为重要。对于传统的协同推荐存在数据处理速度问题,但是已有学者进行了大量研究,并提出了并行计算、近似计算、矩阵分解和增量计算等有效的解决方案。文献[17]和文献[18]可以吸收这些针对大数据的处理的算法和技术,来提高基于二分网络推荐算法处理大数据的能力和效率。

3. 将TAA算法同其他算法融和,进一步提升推荐效果。当前,各种推荐算法融合是推荐算

法设计的趋势,同样,TAA算法也能同其他各种算法进行融合提高推荐效果。

参考文献

- [1] 项亮. 推荐系统实践[M]. 北京:人民邮电出版社, 2012:22-234.
- [2] 张新猛, 蒋盛益, 张倩生, 等. 基于用户偏好加权的混合网络推荐算法[J]. 山东大学学报:理学版, 2015(9):29-35.
- [3] 刘萍, 郑凯伦, 邹德安. 基于LDA模型的科研合作推荐研究[J]. 情报理论与实践, 2015, 38(9):79-85.
- [4] 巩军, 刘鲁. 基于个人知识地图的专家推荐[J]. 管理学报, 2011, 08(9):1365-1371.
- [5] 陈华. 海量学术资源的专家推荐系统分析与设计[D]. 北京工业大学, 2012.
- [6] 李霞, 李守伟. 面向个性化推荐系统的二分网络协同过滤算法研究[J]. 计算机应用研究, 2013, 30(7):1946-1949.
- [7] 许云红. 基于网络方法的专家知识推荐[D]. 合肥:中国科学技术大学, 2010.
- [8] 张斌, 马费成. 科学知识网络中的链路预测研究述评[J]. 中国图书馆学报, 2015(3):99-113.
- [9] Zhou T, Ren J, Medo M, et al. Bipartite network projection and personal recommendation[J]. Physical Review E, 2007, 76(4):70-80.
- [10] Zhang Y C, Blattner M, Yu Y K. Heat conduction process on community networks as a recommendation model[J]. Physical Review Letters, 2008, 99(15):12505-12508.
- [11] Liu J G, Guo Q, Zhang Y C. Information filtering via weighted heat conduction algorithm[J]. Physica A Statistical Mechanics & Its Applications, 2011, 390(12):2414-2420.
- [12] Zhou Y, Liu W, Zhang J. The power of ground user in recommender systems.[J]. Plos One, 2013, 8(8):57-57.
- [13] 韩腾跃. 基于二分网络的个性化推荐系统研究[D]. 南昌航空大学, 2013.

- [14] 贾春晓. 基于复杂网络的推荐算法和合作行为研究[D]. 中国科学技术大学, 2011.
- [15] Zeng A, Chi H Y, Shang M, et al. The reinforcing influence of recommendations on global diversification[J]. *Epl*, 2012, 97(1):18005.
- [16] 吕琳媛. 链路预测[M]. 高等教育出版社, 2013:19-20.
- [17] Hao Ma, Haixuan Yang, Michael R. Lyu, et al. SoRec: Social Recommendation Using Probabilistic Matrix Factorization[C]// Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM 2008, Napa Valley, California, USA, October 26-30, 2008. 2010:931-940.
- [18] Diaz-Aviles E, Drumond L, Schmidt-Thieme L, et al. Real-time top-n recommendation in social streams[C]// ACM Conference on Recommender Systems. ACM, 2012:59-66.