

doi:10.3772/j.issn.2095-915x.2016.04.008

基于二分网络模型的专利权人推荐研究

——以新能源汽车领域为例

周波, 杨朝峰

(中国科学技术信息研究所 北京 100038)

摘要: 通过推荐技术为企业推荐合作者, 在提高技术研发效率方面有着重要的意义。在 5 种 (物质扩散算法、热传导算法、偏热传导算法、混合算法、接受者能力算法) 基于二分网络推荐算法的基础上, 本文提出二阶同向资源扩散算法; 同时使用 β 来判断对合作者合作倾向。以新能源汽车领域专利权人推荐为例进行实验, 实验结果证明使用二阶资源扩散算法比一阶资源扩散算法推荐效果要好, 准确率最高可达 27.59%, 召回率最高可达 30.05%, 提升幅度最大可以达到 15.17%, 最优 β 表明优先选择其曾经有过合作关系的专利权人进行合作。

关键词: 二分网络, 二阶资源扩散, 合作者推荐

中图分类号: G35, TP39

The Recommendation of Patentee Based on Bipartite Network

——A Case Study of New Energy Vehicles

ZHOU Bo, YANG ChaoFeng

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

Abstract: To recommend suitable collaborators through recommendation has important significance of improving the efficiency and the quality of technology research and development. Based on five different bipartite network recommendation algorithms (diffusion algorithm, heat conduction algorithm, partial heat

作者简介: 周波 (1991-), 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-), 副研究员, 博士, 研究方向: 科技政策、宏观经济政策。

conduction algorithm, hybrid algorithm, the receiver ability algorithm) , this paper proposed a new algorithm hereafter, we call it “the second order resource diffusion algorithm of the same direction” , and this study employed a parameter β to evaluate the tendency of cooperation. We applied this new algorithm in the field of new energy vehicles as an example. The experiment results showed that the performances of this new algorithm are better than the first order resource diffusion algorithm, the precision rate is 27.59%, the recall rate is 30.05%, the highest improvement is 15.17% compared with other algorithms. In addition, the results of optimal β indicated that the patentee are preference of former collaborators.

Keywords: Bipartite network, second order resource diffusion, collaborator recommendation

1 引言

在学术界，没有学术合作和研究人员之间合作的科学研究成果是不可想象的。以往的研究已经发现，高产的学者往往是更多合作的结果。寻求合作者是科研工作者的一个重要学术活动之一，因为合适的合作者会有助于提高学者的研究质量，加快其研究进程，提高研发效率。随着网络时代的发展，学术交流越加广泛，合作者推荐技术系统的研究和开发对于拓展学术作者之间的更进一步合作有重大的意义。在过去的十年中，合作者推荐一直是许多研究的焦点。在 Web of Science 数据库中以 $TI=(Collaborator OR Coauthor) AND TI=Recommendation$ 为检索式 (TI 表示主题)，共检索文献 235 篇，其中 2010—2015 年共有 132 篇，近 6 年来文献发表量超过 56%。然而，通过对这些文献研究发现，这些文献大多是针对合著者推荐，进行专利权人推荐的研究是极少，因而进行技术合作中专利权人的推荐是研究的空白。

然而，进行专利权人合作推荐和合著网络中研究者的推荐一样是具有极高的研究价值。随着企业研发的全球化、研发领域的多样化、技术研发规模和难度的加大以及技术的快速变革，寻找合适的技术研发伙伴成为企业越来越强烈的需求；通过推荐技术为技术研发机构推荐高合适的专利

权人，缩短寻找研发伙伴的时间，加快研究进程，提高技术研发效率方面有着重要的意义。

新能源汽车领域是未来汽车工业发展的方向，是工业竞争的前沿，各国政府高度重视，大量的企业正在进行研发，目前新能源汽车还没有普及，表明这个领域的技术研发是存在一定研发瓶颈的，一个企业很难快速研究出成果，中间必定有大量的技术交流和合作研发，这样就有大量的合作关系可以用二分网络来构建。同样正是因为该领域存在较大难度，因此寻找研发合作伙伴就是企业的很大的需求，进行专利权人推荐技术研发合作伙伴可以满足技术研发机构的需求。

2 合作者推荐技术研究文献综述

通过查阅相关文献我们可以发现，当前的合作者推荐主要是用于合著者推荐，而进行专利权人合作推荐的几乎是研究的空白，而且几乎都是基于作者-作者合著网络计算作者之间的相似性，很少使用二分网络进行合作者推荐的相关文献。因此，基于二分网络进行合作者推荐是一片研究空白。

基于二分网络的推荐算法由周涛^[1]2007 基于复杂网络理论而提出，该算法提出后，受到广泛关注，查阅相关文献，目前主要的基于二分网络

的推荐技术主要有以下5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。文献^[2]中,证明了这些算法的结果明显好于经典的协同过滤。

第二种是热传导算法,文献^[3]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。

第三种是偏热传导算法。热传导算法会降低冷门的产品推荐的力度,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^[4]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。

第四种是接受者能力算法。文献^[5]研究了资源接受者的接收能力的差异对推荐精度的影响,提出基于资源接收能力不同的算法。

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,文献^[4]根据物质扩散算法和热传导算法的特点提出了将二者融合的算法。

专利权人-专利合作网络是二分网络的一种,这种网络同客户-商品网络,用户-电影网络和社交网络等不同,这种不同主要体现在以下3个方面:第一,合著网络中的连接是一次性的。例如专利权人 x_1 和 x_2 合著一篇专利 y_1 ,那么作者A,B,C将不会再次合作 y_1 ^[6]。而在客户-商品网络和用户-电影网络等中,用户A和用户B都购买过 y_1 ,用户A,B,C均有可能再次连接到 y_1 。第二,推荐对象不同。在专利权人和专利构成的二分网络中,推荐的是专利权人,是进行同质节点的推荐,而一般的推荐算法是进行异质节点的推荐,例如电子商务中向客户推荐商品,目标对象和推荐对象并不是同一类节点。第三,合

著者之间有很大的可能性再次合作,这同社交网络中的朋友推荐是不一样的。举例如下,在朋友网络中,A和C是朋友,B和C是朋友,则我们可以向A推荐C,而不用向A推荐B,因为他们已经是朋友关系了;但是在合作网络中,例如A和C合作过,A和B合作过,则我们不仅可以向A推荐C,还可以向其推荐B,因为合作的作者之间还有可能再次合作。因此,在同质网络上的基于相似性传递的推荐算法不能适用于合著网络的推荐。

因此,需要重新设计针对专利权人合作者的推荐算法。

3 二阶资源扩散同质节点推荐算法设计

基于以上我们可以总结出以往的基于二分网络推荐算法的路径是资源用户向商品分配,再由商品向用户分配,最后由用户向商品分配。然而要进行专利权人同质节点的推荐,其资源扩散的路径是专利权人到专利,再由专利到专利权人。用资源扩散矩阵来说这是一阶资源扩散。下面将分析使用一阶资源扩散存在的问题。

设有如下的专利权人和专利二分网络,现要对专利权人 x_2 推荐其未来的合作者。使用一阶资源扩散的第一次资源扩散路径如下:

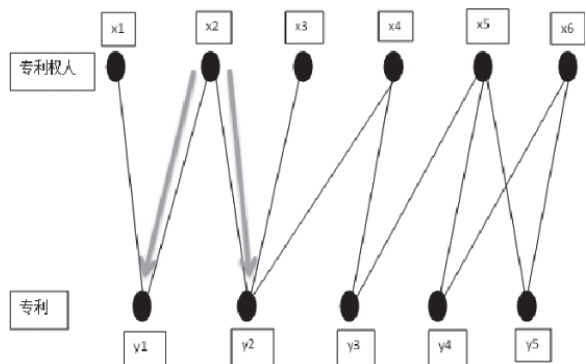


图1 第一次资源扩散路径

第二次资源扩散路径如下：

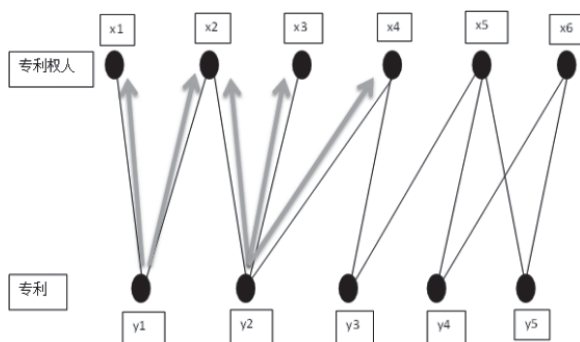


图 2 第二次资源扩散路径

假设忽略资源扩散量的话，专利权人 x_2 的推荐结果为 x_1 和 x_3 ，通过分析 x_1 和 x_3 ，我们可以发现 x_1 和 x_3 都是 x_1 曾经的合作伙伴，未曾合作过的合作者推荐的概率为零。这种推荐是无效的，因为推荐系统是发现用户的隐形信息需求，就是说推荐目标专利权人还没有合过的专利权人，甚至是目标专利权人还不知道的潜在合作者。此外，通过一次物质分配算法，所得的结果直接用于专利权人推荐的话是毫无意义，因为其推荐的都是已经有过合作关系的合作者，对于用户来说是毫无新颖度的。

因此，以往的使用一阶资源扩散的同质节点推荐算法并不能直接用来做专利权人的推荐，需要设计新的推荐算法。

基于此，本文通过考虑二阶资源扩散来构建专利权人的推荐。基本思想就是在一次物质分配结束后再进行一次物质分配，这样可以将资源分配给用户没有过合作关系作者。

二阶资源扩散在一次资源扩散结束后，还要进行两次资源扩散，第三次资源扩散如下为：

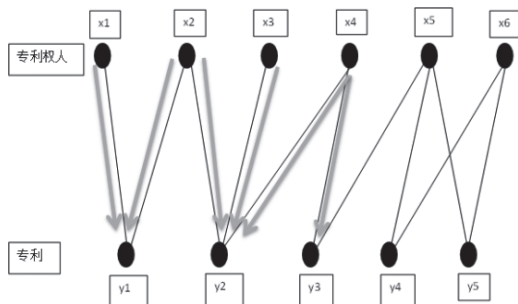


图 3 第三次资源扩散路径

第四次资源扩散路径如下：

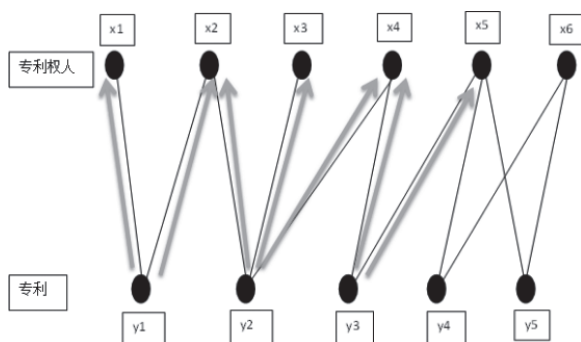


图 4 第四次资源扩散路径

经过二阶资源扩散以后，专利权人 x_2 的推荐合作者有 x_1 、 x_3 、 x_4 和 x_5 。其中 x_1 和 x_3 是 x_2 曾经的合作者， x_4 和 x_5 是新的合作者。

以物质扩散算法为例（热传导算法、偏热传导算法、混合算法、接受者能力算法类同，不再一一列出），用 X, Y 分别表示专利权人和专利的集合，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 X_i 的度和 Y_j 的度。设 X 对物品 Y 的初始资源分配矩阵为 H ，物质扩散算法的一阶资源扩散矩阵 $W1$ ：

$$W1(kj) = \frac{1}{d(Yj)} \sum_{l=1}^m \frac{b(k,l) * b(j,l)}{d(Xl)}$$

则二阶资源扩散矩阵 $W2$ ：

$$W2 = W1 * W1$$

二阶资源扩散形成的推荐列表矩阵为：

$$R2 = W2 * H$$

然而，如上文所提到，专利权人推荐与社交网络的不同，合作者之间可能再次合作，因此我们需要考虑其与有过合作关系和没有过合作关系作者之间发生合作的倾向性问题。尚不清楚其合作倾向的情况下，我们可以做出两种假设。假设一：由于用户已经同有过合作关系的专利权人合作过，彼此之间相互熟悉有合作基础，因此，

用户更加倾向于同有过合作关系的专利权人合作。假设二：用户合作的需求来源于用户自己的研发困难和研发风险，通过合作将风险降低，为了在更大的程度上化解风险，以及为了防范竞争对手和技术来源多元化，用户更倾向于同新的专利权人进行合作。

在无法判断这两种假设的合理性和倾向性程度的情况下，我们引入一个参数来描述这种倾向性程度，在后面的算法中进行具体化。

为了判断合作倾向性问题，需要加入一个参数 β 来表示专利权人的合作倾向性问题， $0 \leq \beta \leq 1$ 。当 β 等于 0.5 时，表示的是新旧合作者之间资源是平等分配的；当 β 等于 1 时，资源只在以往合作过的专利权人之间分配，新的专利权人之间分配的资源数量为 0。当 β 等于 0 时，资源只在新的专利权人之间分配，以往合作过的专利权人之间分配的资源数量为 1。当 $0 \leq \beta \leq 0.5$ 时，目标对象更倾向于新的专利权人之间合作，当 $0.5 \leq \beta \leq 1$ 时，目标对象更倾向于已经合作过的专利权人之间合作。

$$W2'(i, j) = \begin{cases} \beta * W1(i, j), & W1(i, j) > 0 \\ 0, & i = j \\ (1 - \beta) * W2(i, j), & W1(i, j) = 0 \end{cases}$$

$$R2' = W2' * H$$

将目标专利权人的资源量置为零，然后将目标专利权人的推荐结果按照资源量的大小进行排序，取推荐结果前 Top-n，作为最终的推荐列表。推荐结果是随着 β 的变化而变化，寻找最优的推荐结果，确定最优参数 β ，确定专利权人的合作倾向性问题。

4 专利数据处理和推荐算法实验

4.1 专利检索

借鉴文献 [7] 和 [8] 给出了专利检索式确定的一般方法，本文借鉴该文中方法，采用关键词和

IPC 分类号相结合的方法进行设计检索式，弥补单一方法检索的不足。

在关键词的确定上中，本文将参考文献 [9]、[10]、[11] 和 [12] 等文中的关键词方法。由于日本的新能源汽车发展走在世界前列，因此为了更准确地确定关键词，先对丰田、日产与本田公司的介绍中的有关新能源技术措辞进行整理，使关键词符合企业实际。初步确定基本关键词为：混合动力汽车 (Hybrid Vehicle, HV)、电动汽车 (Electric Vehicle, EV) 与燃料电池汽车 (Fuel Cell Vehicle, FCV)。然后通过百科，字典，论文查找关键词的同义词，近义词和相关词，其中混合动力汽车的同义词还有 Hybrid Electric Vehicle, HEV, Hybrid Electric Car。由于新能源汽车与传统汽车不同的关键点在于电池技术，因此将电池技术也设为基本关键词。

对 electric vehicle 这一关键词检索时发现有很多汽车电子控制系统或电子零件等非电动汽车专利信息，需要将这部分专利剔除。因此需要设计出更加具体检索式，根据得到的专利摘要进行分析，确定指向电动汽车专利的关键词：electric car、electric vehicle、electric motor car、electric motor vehicle、electrically driven vehicle、electrical driven vehicle、electric automobile、electric traveling vehicle、electric motorcar、electrical powered vehicle、charge controller for vehicle、electric storage device and vehicle、battery pack^[13]。与电池技术相关的关键词是 battery, fuel cell, FC, fuel cell stack, motor, electric motor, control, current sensor, charge, gear system, inverter^[14]。根据关键词进行主题检索，得出如下检索式：

得出电动汽车的汽车检索式一：TS=(Electric Vehicle OR EV OR electric car OR electric vehicle OR electric motor car OR electric motor vehicle OR electrically driven vehicle OR electrical driven vehicle OR electric automobile OR electric traveling vehicle

OR electric motorcar OR electrical powered vehicle
OR charge controller for vehicle OR electric storage
device and vehicle OR battery pack)

得出混合动力汽车专利的检索式二：
TS=(Hybrid Vehicle OR HV OR Hybrid Electric
Vehicle OR HEV OR Hybrid Electric Car)

得出混合动力汽车专利的检索式三：
TS=(Fuel Cell Vehicle OR FCV)

得出电池技术专利的检索式四 :TS= (battery
OR fuel cell OR FC OR fuel cell stack OR motor
OR electric motor OR control OR current sensor
OR charge OR gear system OR inverter)

最终确定的专利关键词检索式五：(检索式
一 OR 检索式二 OR 检索式三) AND 检索式五

通过查找中文版《国际专利分类表关键词索引》和文献^[6]，根据检索的主题和关键词检索IPC,查找到以下的IPC分类号同新能源汽车相关。相关的IPC分类号如下表1。

表 1 IPC 分类号表

技术类别	技术领域	IPC 分类号
电池技术	电动装置的布置或安装	B60K1
	电路或流体管路及其原件布置	B60R16
	电极和电极件	H01M4
	结构零件或制造方法	H01M2
燃料电池汽车	电池及其制造	H01M10
	机动车; 挂车	B62D
	电源的电力牵引	B60L11
	燃料电池制造技术	H01M8
纯电动汽车	车辆传动装置的布置与安装	B60K17
	电池的充电装置	H01M2
	紧急保护电路装置	H02H
	电机	H02K
混合动力汽车	供电或配电的电路装置或系统, 电能存储系统	H02J
	原动机的布置或安装	B60K6
	车辆的控制系統	B60W20
	车辆子系统的联合控制	B60W10

分析文献 [15]、[16] 和 [17] 的检索式，在电动汽车的专利检索结果中有部分是游戏玩具汽车和一些普通汽车的专利，这部分专利是要剔除掉的，这两类的IPC分类号分别是：B62(无轨路用车辆)，A63H(玩具，如陀螺、玩偶、滚铁环、积木)。

因此，最终确定的IPC分类号的专利检索式六：IP= ((B60* OR H01M OR H02H OR H02J OR H02K OR B62D) NOT (B60J OR B60S OR B60V OR B60R1/* OR B62* OR A63H))

结合关键词和IPC分类号的检索式确定最终的检索式为：检索式五和检索式六。

在德温特专利数据库组织检索，检索时间为2016年1月18号，时间跨度为1990-2016年。检索结果共86804件专利。

专利数据清洗。第一步：将Web of Science上的数据下载导入Excel中，只保留PN(专利号)和AU(专利权人)这两个字段。第二步：提取专利权人代码。在德温特专利数据库标准代码中子公司和母公司是使用同一个专利权人代码，使用专利权人代码将子公司和母公司统一进行考虑。这样做的好处是，在推荐结果中避免出现向子公司推荐母公司或像母公司推荐子公司，因为子公司之间相对比较熟悉和了解，对于需要推荐服务的企业来说意义不大，使用专利权人编号进行推荐，推荐的结果价值更大。当然，对于被非标准编码的公司中子公司和母公司就不一定是使用同一个专利权人代码，但是由于在检索结果中，非标准代码表示的公司较少，不到5%因此忽略不计。第三步：由于计算能力有限，删除非合作专利及有5个专利权人以下的合作专利。第四步：生成五组随机数。20%的数据作为测试数据，80%作为训练数据。

4.2 实验结果评价指标

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

	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：

$$\text{recall} = a/a+c$$

算法伪代码如下：

第一步：数据清洗，生成专利权人—专利矩阵。

第二步：计算专利权人和专利节点的度。

第三步：进行一阶资源扩散，得到一阶资源扩散矩阵 W1。

第四步：进行二阶资源扩散，得到二阶资源扩散矩阵 W2，再根据合作倾向，进行资源重新分配，得到二阶资源分配矩阵 W2'。

第五步：二阶资源扩散形成的推荐列表，截取前 top-n 作为推荐结果。

第六步：计算查全率和查准率对推荐结果进行评价。

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。

5 实验结果分析与讨论

5.1 实验结果分析

各种算法推荐结果如图 5、图 6、图 7、图 8

和图 9 所示 (其中，偏热传导算法、混合算法、接受者能力算法的推荐结果在 top-20、top-30、top-40 和 top-50 的趋势同 top-10 类似，不再一一展示)，可以发现五种算法推荐结果的准确率随着推荐列表的长度增加而降低，而召回率随着推荐列表的长度增加而提高，这规律是同信息检索领域的查准率和查全率之间的关系是一致的^[18]。推荐结果的召回率和准确率都是随着合作参数 β 的增加先增后减，由此证明了本文所提的二阶资源扩散算法的正确性，对实验结果数据分析可以确定各个算法的最优合作参数 β 。表 3 是各个算法在不同的推荐长度下的最优合作倾向参数 β 值。不同的算法的最优合作参数不同的，而同一算法不同的推荐长度对应的最优合作倾向参数也是不同。但是所有的 β 值为 $0.5 \leq \beta < 1$ ，说明目标对象在选择合作者的时候确实存在合作倾向的问题：目标对象更倾向于同已有过合作经历的专利权人合作，但是并不是完全只和已有过合作经历的专利权人合作，还在较小的程度上同没有合作过的专利权人合作。

表3 各算法在不同推荐长度下的最优 β 值

推荐列表长度	物质扩散算法	热传导算法	偏热传导算法	融合算法	接受者能力算法
top-10	0.7、0.8、0.9	0.9	0.8	0.7、0.8、0.9	0.7、0.8、0.9
top-20	0.8、0.9	0.9	0.9	0.8、0.9	0.8、0.9
top-30	0.5	0.9	0.9	0.5	0.5
top-40	0.5	0.9	0.9	0.5	0.5
top-50	0.5	0.9	0.9	0.5	0.5

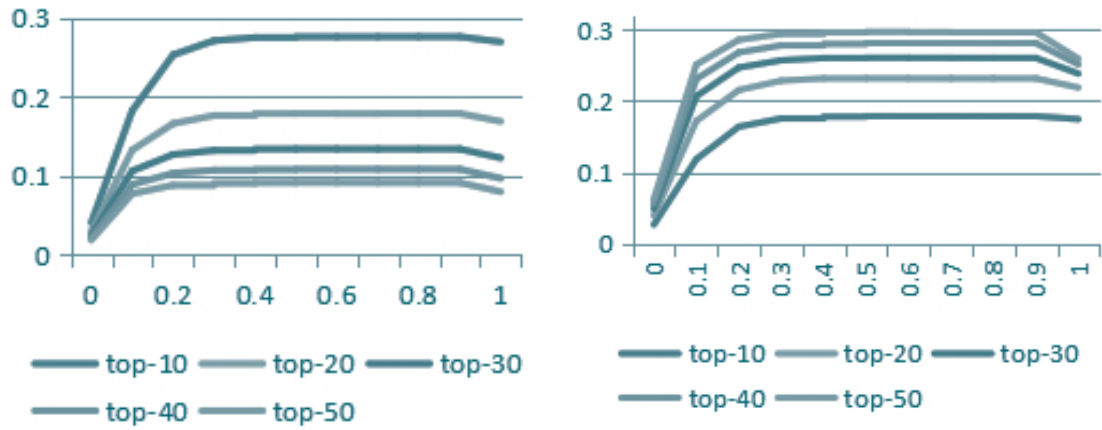


图5 物质扩散算法的准确率(左)和召回率(右)随着 β 值变化情况

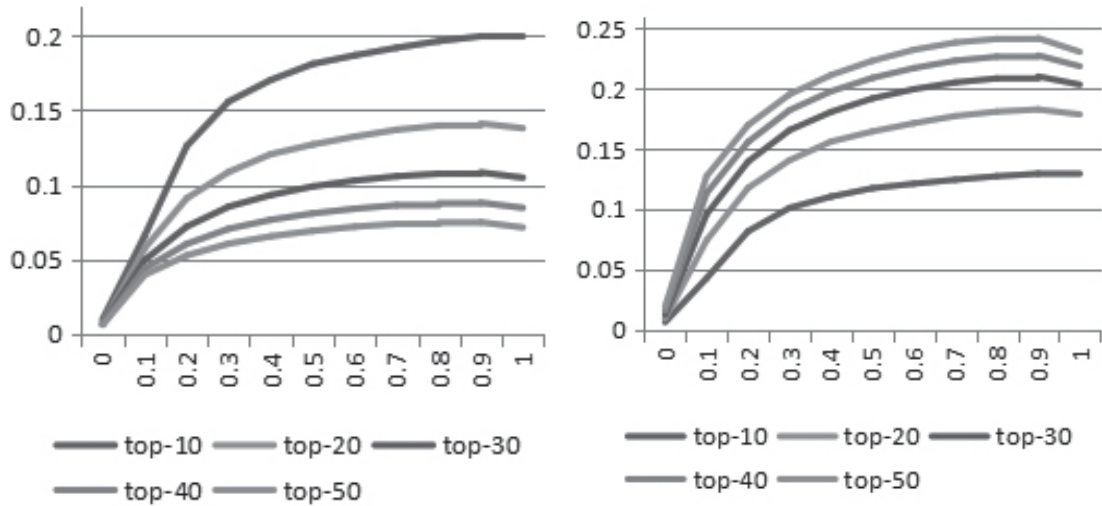


图6 热传导算法的准确率(左)和召回率(右)随着 β 值变化

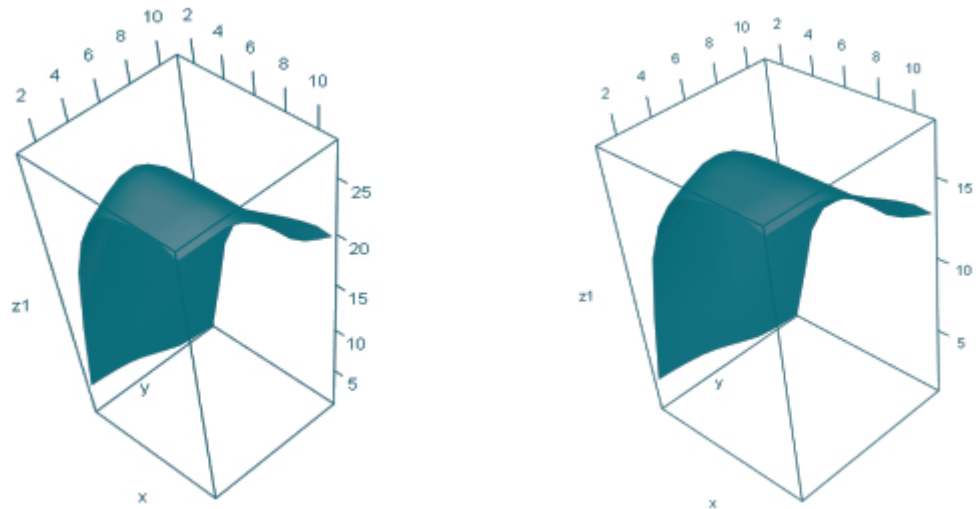


图7 在 top-10 时偏热传导算法的准确率(左)和召回率(右)随着 β (x轴)和 λ (y轴)值变化

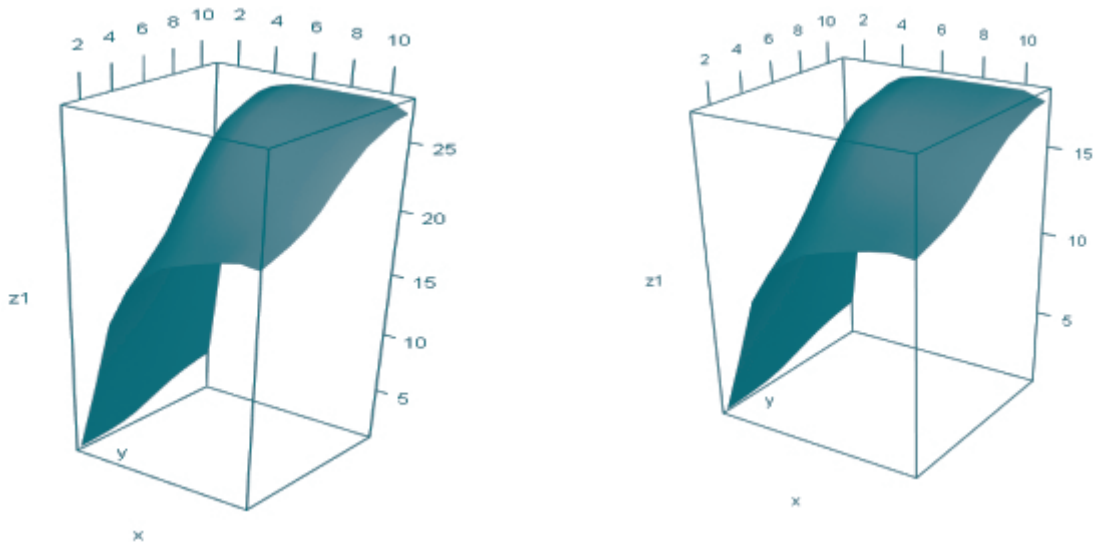


图8 在 top-10 时混合算法的准确率 (左) 和召回率 (右) 随着 β (x 轴) 和 λ_2 (y 轴) 值变化

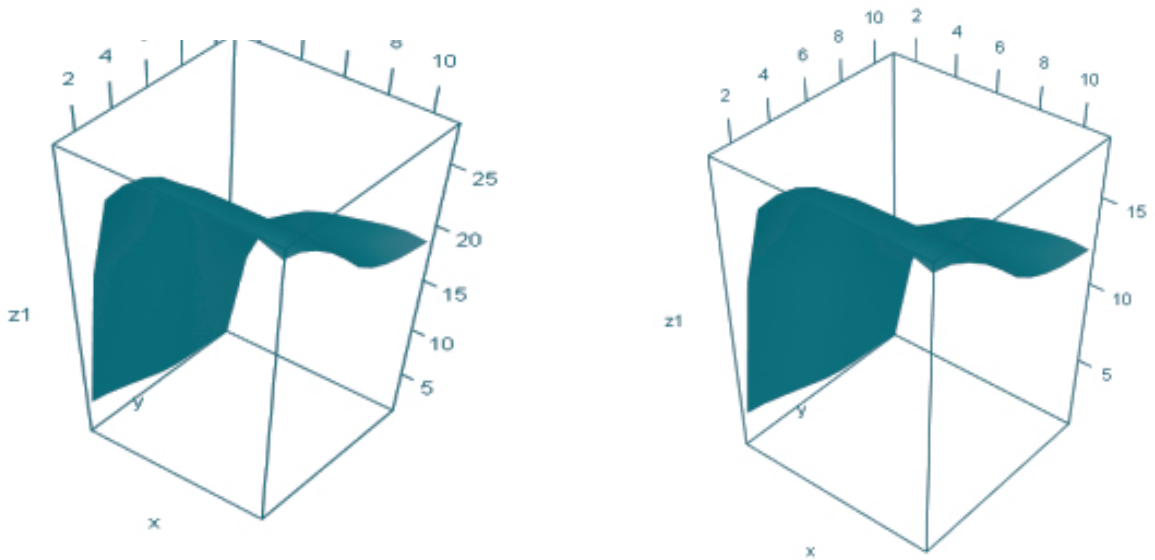


图9 在 top-10 时接受者能力算法的准确率 (左) 和召回率 (右) 随着 β (x 轴) 和 λ_3 (y 轴) 值变化

偏热传导算法、融合算法和接受者能力算法分别需要确定最优参数 λ_1 、 λ_2 、 λ_3 才能确定最优的推荐结果, 图 10 是偏热传导算法在最优合作倾向参数下, 推荐结果随着 λ_1 ($0 \leq \lambda_1 \leq 1$) 的

变化情况。图 11 是融合算法在最优合作倾向参数下, 推荐结果随着 λ_2 ($0 \leq \lambda_2 \leq 1$) 的变化情况。图 12 是接受者能力算法在最优合作倾向参数下, 推荐结果随着 λ_3 ($0 \leq \lambda_3 \leq 1$) 的变化情况。

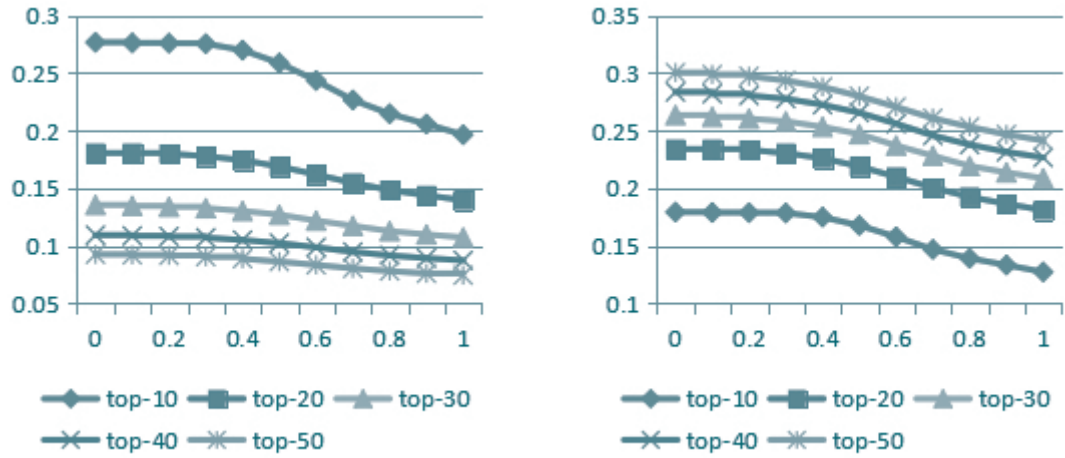


图 10 在 $\beta = 0.8$ 时，偏热传导算法的推荐的准确率（左）和召回率（右）随着 λ_1 的变化

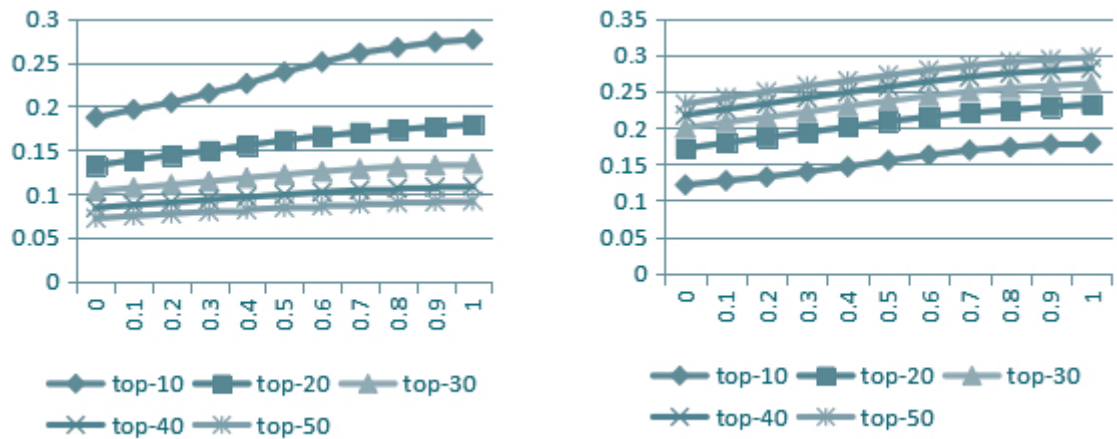


图 11 在 $\beta = 0.6$ 时，融合算法的推荐的准确率（左）和召回率（右）随着 λ_2 的变化

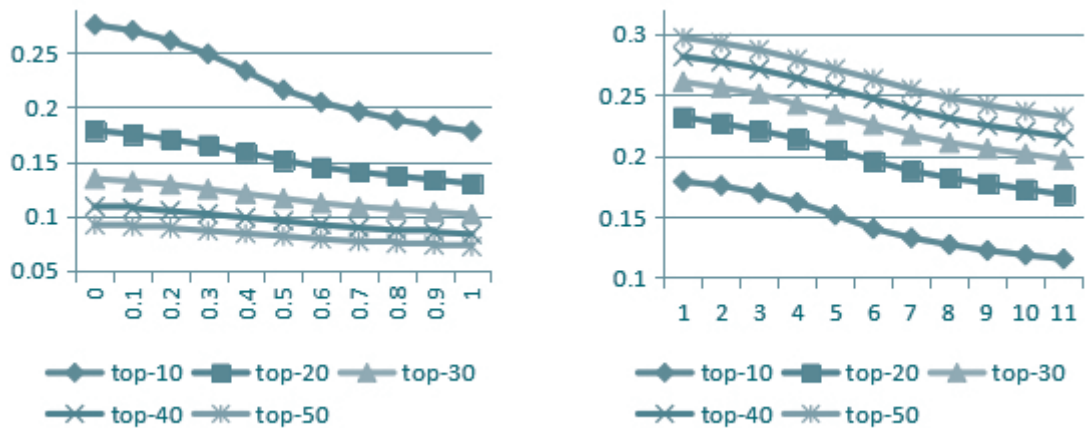


图 12 在 $\beta = 0.6$ 时，接受者能力算法的推荐准确率（左）和召回率（右）随着 λ_3 的变化

通过图 12 可以发现当 $\lambda_1=0$ 、 $\lambda_2=1$ 、 $\lambda_3=0$ 时，推荐结果的准确率和召回率最大。然而由于设定的 λ_1 、 λ_2 、 λ_3 的范围均为 0 到 1 之间，无法确定 $\lambda_1=0$ 、 $\lambda_2=1$ 、 $\lambda_3=0$ 时推荐结果是否为最优。因此，为了确定最优参数还需进一步实验。通过分析可以发现，随着推荐长度的增加，推荐结果的趋势基本一致，故选择仅在推荐列表长度为 10 时进行实验。偏热传导

算法的参数范围 $-1 \leq \lambda_2 \leq 0$ ，混合算法的参数范围 $1 \leq \lambda_2 \leq 2$ ，接受者能力算法的参数范围 $-1 \leq \lambda_3 \leq 0$ 。实验结果如图 13，可以得出这三种算法的最优参数分别为 $\lambda_1=0$ 、 $\lambda_2=1$ 、 $\lambda_3=0$ 。通过对比可以发现，当 $\lambda_2=1$ 、 $\lambda_3=0$ 就是物质扩散算法。因此，仅需对物质扩散算法、热传导算法、偏热传导算法的推荐结果进行对比，如图 14。

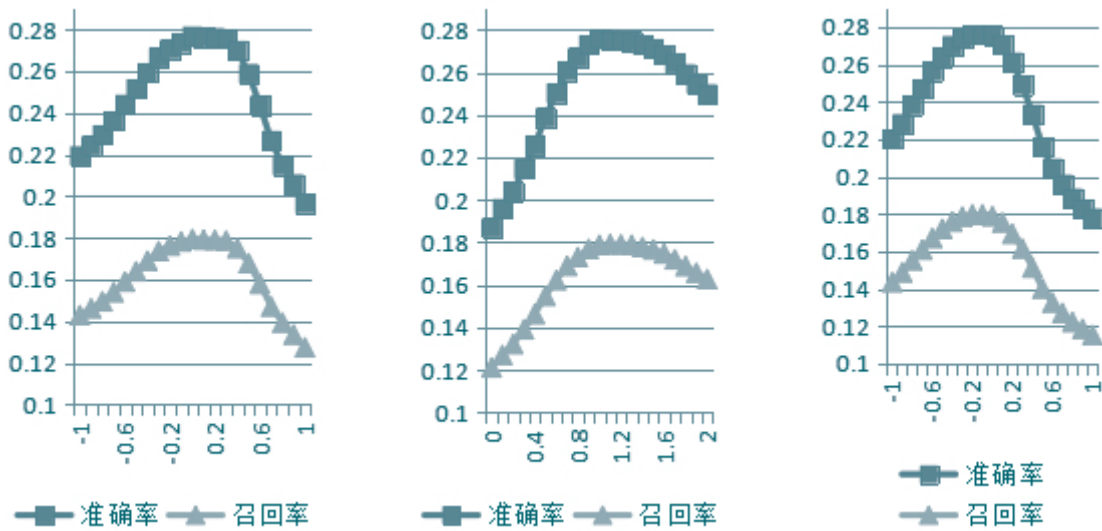


图 13 在 top-10 时，偏热传导算法 $\beta=0.8$ 时推荐结果随着 λ_1 的变化 (左)、融合算法 $\beta=0.6$ 时推荐结果随着 λ_2 的变化 (中)、接受者能力算法 $\beta=0.6$ 时推荐结果随着 λ_3 的变化 (右)

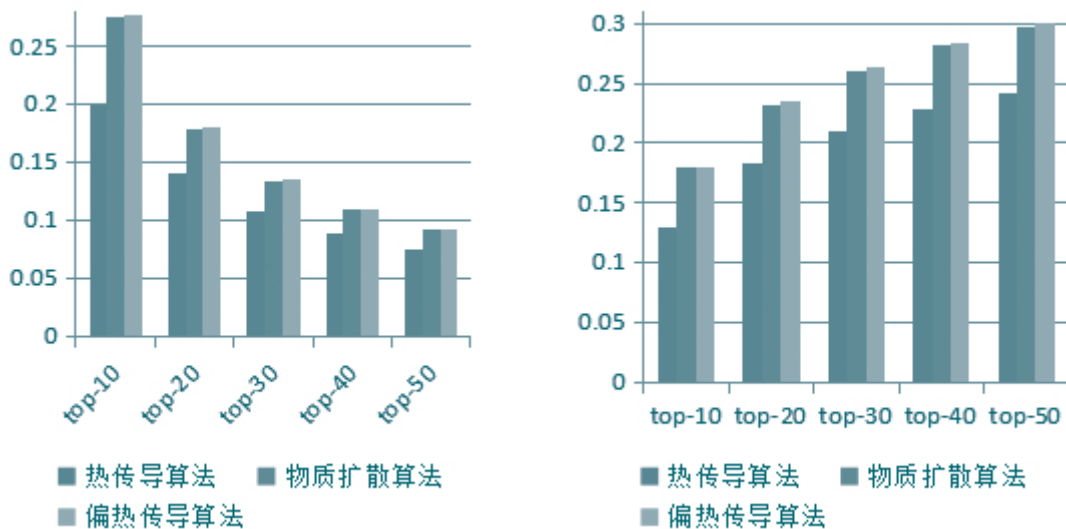


图 14 在最优参数的情况下，各种算法的推荐算法准确率 (左) 和召回率 (右) 比较

将二阶资源扩散推荐算法同一阶资源扩散推荐算法推荐结果进行对比,如图 15 所示。可以发现采用二阶资源扩散的推荐结果比一阶资源扩散的结果要好,并且随着推荐列表长度的增加,二阶资源扩散算法推荐效果的提升度也在增加,在 top-50 的推荐效果提升度最高可达到 15.17%,提升效果明显。

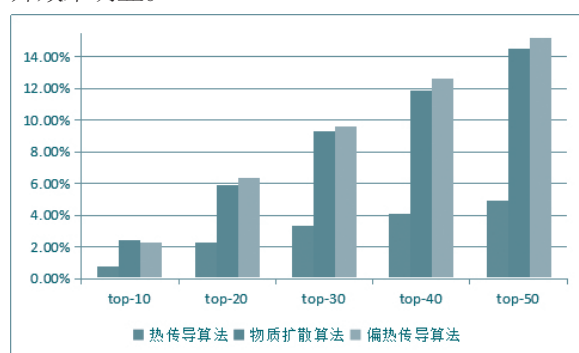


图 15 同一阶资源扩散算法相比,采用二阶资源扩散算法推荐结果的提升度

5.2 结果讨论

通过实验结果可以发现,偏热传导算法推荐效果最优,物质扩散算法推荐效果要略低于偏热传导算法,但是差距不超出 0.3%,混合算法和接受者能力算法在最优参数取值的情况下,实质就是物质扩散算法.热传导的算法推荐效果最差,这同各个算法所适用的推荐对象不同,不同的网络之间合作关系形成机理不同。物质扩散算法主要适用推荐热门对象,而热传导算法适用于推荐冷门对象^[9]。在推荐合作者时可以这样理解:物质扩散算法更倾向于推荐熟悉的专利权人,而热传导算法更倾向于推荐不熟悉的专利权人,这同上文的实验结果相吻合:在上文已通过实验证明,合作倾向参数 β 值范围 $0.5 \leq \beta < 1$,即专利权人之间更倾向与有个合作关系的专利权人合作,而不倾向与不熟悉的专利权人之间合作研发。这同专利权人的合作网络形成机理相同,专利权人优先选择熟悉的合作者,其次再同新的合作者进

行合作,故偏向冷门推荐的热传导算法推荐的效果较差。因此,物质扩散的算法要好于热传导算法,混合算法中热传导算法,在最优参数的情况下,出现热传导算法无效,只是物质扩散算法起到推荐作用。同样的理由也可以解释接受者能力算法和偏热传导算法最优参数 $\lambda_1=0$ 、 $\lambda_3=0$ 的现象。其实,接受者能力算法和偏热传导算法 λ_1 、 λ_3 主要是用于调节对冷门对象的推荐力度,当 $\lambda_1=0$ 、 $\lambda_3=0$ 是对冷门对象的推荐效果为 0,大于 0 或小于 0 都会加大对冷门对象的推荐力度,会降低推荐效果。

6 总结

本文对于合作者推荐,提出了基于二分网络的二阶资源扩散算法。实验结果证明使用二阶资源扩散算法比一阶资源扩散算法的推荐效果要好,提升幅度最大可以达到 15.17%,并且提升幅度随着推荐列表长度的增加而增加。对于专利权人合作的倾向问题,提出了两个假设并在推荐算法中设置合作倾向参数,通过实验检验假设的正确性,证明专利权人更倾向与有过合作关系的专利权人合作。对 5 种有代表性的二分网络推荐算法进行了实验,实验结果表明,物质扩散算法和偏热传导算法效果较好,准确率最高可达 27.59%,召回率最高可达 30.05%。并对各种推荐算法的推荐结果做出了解释。

但是由于各种客观原因,本文也存在诸多纰漏,有待进一步的研究。关于检索式的权威性,由于缺乏公开的专利权人合作关系数据集,本文选择以新能源汽车领域的专利权人合作为例来进行研究。进行专利检索时,尽管参考了大量的文献,但是毕竟对新能源汽车领域不熟,其权威性可能存在不足。关于使用新能源汽车领域这一个领域进行实验可能存在的代表性问题,本文所提

的基于二分网络的二阶资源扩散算法进行专利权利人推荐,只是选择了新能源汽车领域来进行实验,还有待于使用更多不同的数据集进行验证。删除部分数据的合理性问题,受计算能力的限制和数据稀疏性问题,本文清除了4个及4个合作者以下的专利,合作关系不足一万对,数据量有限,算法的是否有效性还有待于进一步验证。

未来的研究可在本文的基础上,采用更多的数据集、更多的技术领域来进行进一步深入研究,以弥补本研究的不足。

7 参考文献

[1]Zhou T, Ren J, Medo M, et al. Bipartite Network Projection and Personal Recommendation[J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2007, 76(2):70-80.

[2]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):12505-125.

[3]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.

[4]韩腾跃. 基于二分网络的个性化推荐系统研究[D]. 南昌:南昌航空大学, 2013.

[5]贾春晓. 基于复杂网络的推荐算法和合作行为研究[D]. 合肥:中国科学技术大学, 2011.

[6]吕林媛,周涛. 链路预测[M]. 北京:高等教育出版社, 2013.

[7]罗晓宁,郑乃章. 科技查新中专利检索式的构建[J]. 科技情报开发与经济, 2014(18):112-114.

[8]庞德良,刘兆国. 基于专利分析的日本新能源汽车技术发展趋势研究[J]. 情报杂志, 2014(5):60-65.

[9]李纲,任佳佳,毛进,等. 专利权利人合作网络的社群结构分析——以燃料电池电动汽车专利为例[J]. 情报学报, 2014, 33(3):267-276.

[10]王智琦,陈悦,姜照华,等. 前沿技术与核心技术识别的投入产出分析方法——以混合动力汽车为例[J]. 科学学研究, 2015, 33(11):1612-1620.

[11]高幸,柴华奇,汤松. 基于专利引文数据的混合动力汽车创新扩散研究[J]. 世界科技研究与发展, 2013(6):784-787.

[12]吴为理,黄维军. 基于专利分析的混合动力汽车技术分布研究[J]. 公路与汽运, 2011(6):5-9.

[13]黄远辉. 基于专利地图的混合动力电动汽车竞争情报研究[J]. 现代情报, 2010, 30(11):53-58.

[14]罗璇. 基于数据挖掘技术的专利信息分析及应用研究[D]. 北京:首都经济贸易大学, 2011.

[15]李伟,刘红光. 国外混合动力汽车领域专利引证分析[J]. 情报杂志, 2011, 30(9):6-13.

[16]徐国祥,余碧涛,李福燊,等. 我国新能源汽车领域专利申请情况分析[J]. 电池, 2012, 42(5):289-292.

[17]栾春娟. 全球燃料电池领域技术网络结构的演进[C] // 第九届中国科技政策与管理学术年会论文集, 2013:1-8.

[18]许忠锡. 关于查全率与查准率相互关系问题的新认识[J]. 情报杂志, 2004, 23(3):57-60.

[19]王金燕. 推荐系统中标签推荐的研究[D]. 南京:东南大学, 2012.