

基于社会网络的学科主题聚类研究*

朱梦娴 程齐凯 陆伟

(武汉大学信息资源研究中心 武汉 430072)

摘要 将社会网络分析的方法运用到学科主题聚类中,对从CNKI上采集的发文数据构建关键词共词网络,再用Blondel社区发现算法进行关键词聚类分析,接着用Z-value对划分的社区进行核心圈分析。最后,笔者选取三个适当的例子对聚类算法和核心圈的结果进行评价,得到:Blondel社区发现算法虽然存在一些缺陷,但在数量和质量上都效果不错;Z-value确定社区核心圈效果很好,核心节点能很好地表征社区内容。在此基础上,笔者得出了图书情报学领域的主题聚类结果。

关键词 学科主题聚类 社会网络 社区发现 Z-value 核心圈

中图分类号 G353.1

文献标识码 A

文章编号 1002-1965(2012)11-0040-05

A Clustering Study of Subject Theme Based on Social Network

Zhu Mengxian Cheng Qikai Lu Wei

(Information Resource Research Center, Wuhan University, Wuhan 430072)

Abstract In this paper, the authors applied the method of social network analysis to the subject theme clustering. The authors built a co-words network based on the 3244 articles collected from CNKI, then clustered the keywords using the Blondel community detecting algorithm, and finally used the Z-value to detect the core circle of the community. At the end, the three selected appropriate examples were used to evaluate the two algorithms, the results showed that the "Blondel community detecting algorithm" runs very good although there exists some problems, and the "Z-value" algorithm runs perfectly good in every aspects. According to these methods, the subject theme clustering of Library and Information Science was concluded.

Key words clustering of subject theme social network community detecting Z-value core circle

聚类作为一种组织信息的有效机制,可以帮助人们更好地组织和利用信息。它在进行用户兴趣分析、发现潜在主题、信息检索等方面都有许多实际的应用。在各学科发展延伸的过程中,利用聚类技术进行学科主题的相关研究较为常见。有效地对学科主题进行聚类,以探测学科研究的热点、划分出学科的不同研究方向,研究其演化方向,是非常必要的。

学科主题聚类是指通过对聚类对象进行主题分析和提取,将聚类对象转换为基于主题的表达形式,本文将社会网络分析方法中的共词分析方法引入学科主题聚类中,基于词与词的共现关系和Blondel社区发现算法来对学科主题进行聚类,并对聚类效果进行评价,方法上相对单纯的聚类方法有所创新,对学科主题的

聚类结果也从侧面反映了所选学科在近年来的研究热点和动向,为今后该学科的研究提供了一定的参考价值。

1 国内外研究现状

相关研究存在于社区划分和主题聚类两个方面。社区划分研究方面:社区划分算法主要有两种思想,分别是凝聚法和分裂法。2004年,Newman等人首先提出发现和评价社区结构的方法^[1],Latapy和Pons提出用随机游走模型在大的网络中计算社区结构^[2],而Palla等人则将社区发现技术应用于社会学和自然科学领域中^[3]。2008年Blondel等人在Newman快速算法的基础上提出了Blondel社区发现算法^[4],其基

收稿日期:2012-06-02

修回日期:2012-08-05

基金项目:教育部人文社会科学基地重大项目“面向细粒度的网络信息检索模型及框架构建研究”(编号:10JJD630014);国家自然科学基金面上项目“基于语言模型的通用实体检索建模及框架实现研究”(编号:71173164)的研究成果之一。

作者简介:朱梦娴(1989-),女,硕士研究生,研究方向:信息资源管理;程齐凯(1989-),男,博士研究生,研究方向:信息检索;陆伟(1974-),男,博士,教授,研究方向:信息检索、知识管理。

本思想与 Newman 快速算法一致,通过迭代的方法划分社区层次。2008 年 Rosvall 和 Bergstrom 等人提出了“理论信息聚类方法”,将复杂网络中大型社区节点进行聚类,但此方法只适用于加权有向网络^[5]。2010 年 Martin 等人在此方法的基础上提出了适用于所有网络形式和聚类算法的方法——“重要性聚类方法”,并将其与 Bootstrap 等方法结合,研究大型网络中社区的演化过程^[6]。

学科主题聚类的研究方面:林颂坚^[7]和张晗^[8]等人分别对计算机领域和生物信息学领域文献的高频词进行聚类,得到各领域的研究热点。2005 年,黄小燕统计五年内情报学领域论文的关键词词频,归纳出五年来情报学领域的研究热点及其演化^[9]。2006 年,马费成和张勤同样利用词频分析方法,分析了国内外知识管理的研究热点^[10]。2007 年,屈鹏、隆捷等人对 LI-SA 中 1996-2003 年的数据进行分析,使用等级聚类方法,得出国际情报学发展的变化轨迹与研究现状^[11]。2008 年,王翼、杜楠、吴斌利用 John Hopcroft 方法对中国生命科学中 1993-1997 年共 150 万篇期刊论文进行社团发现,找出国内医学子领域的研究热点^[12]。同年章成志等人提出基于主题聚类的学科热点及其趋势检测方法,结果表明效果较好^[13]。

在国内外社会网络的研究几乎与主题聚类研究同步发展,在研究主题聚类的过程中很多学者也运用到了社会网络的相关方法,进行社团发现,完成主题聚类,但这类研究还不够完善,有待进一步发展。

2 实验研究方法

2.1 Blondel 社区划分算法 社区是网络科学中的一个重要概念,社区意味着这样的一些节点的集合:节点同社区内节点联系非常紧密,而同社区外节点联系足够少。社区发现的实质是聚类^[14]。应用社区发现,可以找到相邻的节点集,从而发掘出网络的结构特征。目前,已经出现了很多社区发现方法,但是大多数方法都不能处理规模较大的网络,2008 年 Vincent D. Blondel 在 Newman 快速算法^[15]的基础上提出了一个简单的方法来探测大数据网络中的社区结构,它是建立在模块度优化算法上。Blondel 等人的实验结果表明,在计算时间上,这种算法优于任何一种现有的社区发现算法。本文试图用此 Blondel 社区划分算法,从新的角度去尝试对学科主题进行聚类。

模块度的计算公式如下^[16]:

$$Q = \frac{1}{2m} \sum [A_{ij} - \frac{k_i k_j}{2m}] \delta(c_i, c_j) \quad (1)$$

公式中 A_{ij} 代表节点 i 和 j 之间连线的权重; $k_i = \sum_j A_{ij}$ 代表连接节点 i 的所有边的权重之和; c_i 是节点

i 所属的社区;当 i 和 j 属于同一社区时 $\delta(c_i, c_j) = 1$, 否则则为 0; $m = 0.5 \sum_{ij} A_{ij}$ 。

这种算法的思想是:首先将网络中的每个节点看成是一个独立的社区,慢慢将邻近的节点合并,如果合并之后整个网络的模块度提高,那么就合并,否则撤销;如此循环,直到网络的模块度无法提高为止;接着再把每个社区当成一个节点,对每个社区进行如此的合并算法,直到整个网络的模块度无法提高为止。算法描述如下:

a. 将集合中的每一个节点看成一个单独的社区,计算整个网络的模块度。

b. 将任意一个节点 i 的邻接节点 j 合并到 i 所在的社区中,重新计算模块度,如果模块度有所提高,则完成合并;反之若模块度不变或下降,则将 j 放回原来的社区中。

c. 重复步骤 b,直到整个网络的模块度不再提高为止。

d. 将步骤 c 后得到的各个社区作为节点,重复 a—c 的步骤。

e. 重复 a—d 的步骤,直到无论如何合并,整个网络的模块度都不再提高为止。

按照以上迭代的思想,当步骤 c 完成后,可以得到一个网络模块度最高的社区,整个步骤完成后,可以得到多层次的网络社区划分结果。

在研究中,笔者将运用这种算法对构建的网络进行社区划分,将整个网络中的关键词聚成不同层级的类别,并加以分析。

2.2 基于 Z -value 的社区核心圈发现算法 网络社区常常存在着核心节点,它们在网络中的连通性非常好,一个社区中因为有了这些节点,才使得其他节点能够更好地互相联系,一旦失去了这些节点,整个社区的网络结构就会遭到破坏。核心节点组合在一起就叫做社区中的核心圈。A. Mislove 等认为网络的结构跟网络的少数核心高度相关,这些核心的移去将导致图的连通性变得极差,这样,利用节点去除前后的网络连通性的变化就可以测度节点的核心度^[16]。Mislove 的方法计算复杂度较高,在大规模网络中并不适用。R. Guimera 等人提出了一种节点测度核心度的 Z -value 指标用于衡量模块 T 中的节点 i 与模块中其他节点的连接性是否良好。 Z -value 的计算公式如下^[17]:

$$Z_i = \frac{k_{s_i}^i - \langle k_{s_i}^j \rangle_{j \in s_i}}{\sqrt{\langle (k_{s_i}^j)^2 \rangle_{j \in s_i} - \langle k_{s_i}^j \rangle_{j \in s_i}^2}} \quad (2)$$

公式中,代表在模块 s 中节点 i 的度数 s_i 即代表节点 i 所属的模块,平均值 $\langle \dots \rangle_{j \in s}$ 将模块 s 中的所有节点都考虑进去了。

根据经验,在一个社区中,Z-value 值大于 2.5 的节点为核心节点,小于 2.5 的节点为非核心节点,核心节点的连通性好,常常被用来代表社区。

在本文的研究中,笔者将聚类得到的每个社区中每个节点的 Z-value 按照公式计算出来,从高到低排序,将 Z-value 大于 2.5 的节点提取出来作为社区的核心节点,分析每个社区中的核心节点是否能表征整个社区的主要研究内容。

3 数据来源

笔者选取了 CNKI 上图书情报领域的两种权威期刊《中国图书馆学报》和《情报学报》近 12 年的数据作为研究对象,试图分析图书情报领域的学科主题聚类效果,共计发文总数 3 589 篇。设 D 为所有发文数据,则 $D = \{D_1, D_2, D_3, \dots, D_n\}$,其中 D_i 代表每篇文章, $D_i = \{ID, cntitle, cnauthor, address, journal, year, cnkeywords, cnabstract\}$,每个字段分别表示编号、文章名、作者名、作者单位、期刊名、发表年份、中文关键词、中文摘要。笔者将这些数据导入数据库中,以便于分析。

我们去除了征稿启事、会议召开、年度目录等类型的文章,得到 3 244 篇论文作为研究对象。统计显示,这 3 244 篇文章共包含 6 608 个唯一关键词,它们共出现了 13 099 次,这意味着平均每篇文章的关键词为 4. 038 个。这 6 608 个关键词中存在泛词以及同义异形等现象,会影响聚类效果。故笔者对关键词进行了规范化处理:首先,建立一个泛词表,包含“方法”、“希望”、“问题”、“研究”等泛词,无法表征文章的内容,将之删除;其次,由于主题词是规范的、专指度高的关键词^[18],应当作为关键词标引的标准,笔者对照《分类主题词表》将图书情报学领域的主题词查找出来;最后,以主题词为标准,进行关键词的同义异形处理,并将表示同种含义的词语统一化。经过关键词的规范化处理,剩余 6 243 个关键词。

4 实验过程

4.1 构建共词网络 笔者对 1999 - 2010 年《中国图书馆学报》和《情报学报》的 6 243 个关键词构建一个带权重的无向共词网络。在得到网络的共词数据之后,选用 UCINET 工具进行绘图。

4.2 关键词聚类 笔者运用 Blondel 社区发现算法对上述构建的共词网络进行关键词聚类,将整个网络分成了五个层次不同的社区划分结构。表 1 列出了网络中社区划分后词频数排在前 10 的 10 个节点的所

属社区结果。表中 level 表示社区划分的层次,level_0 表示按照 Blondel 算法第一次划分的结果,level_1 表示第一次划分结束后,将每个社区当成节点进行第二次划分的结果,level_2、level_3、level_4 分别表示第三、四、五次划分的结果。

表 1 前 10 个节点的社区划分结果

id	label	level_0	level_1	level_2	level_3	level_4
169	数字图书馆	6	24	201	196	196
33	图书馆	14	32	86	84	84
72	信息服务	14	32	86	84	84
615	知识管理	74	20	164	159	159
201	信息检索	15	1	96	94	94
35	竞争情报	16	305	197	192	192
129	信息资源	14	32	86	84	84
263	情报学	43	6	65	63	63
340	图书馆学	43	6	65	63	63
3792	公共图书馆	501	85	57	200	200
.....						

从表 1 中可以看到,进行第五次划分得到的 level_4 和 level_3 无任何区别,这是因为当确保整个网络的模块度无法继续提高,即无法对任何社区进行合并时,Blondel 算法就会停止运行;可以看到,节点 32、72 和 129 在第一次划分后分别一直属于同一社区,在全部的数据集中,有很多数据一开始分属不同的社区,到最后归于一个社区中。

表 2 列出了三个层次的社区数量情况,具体聚类效果在 5.1 节中将做详细分析。

表 2 五个网络层级分别包含的社区数量

网络层级	level_0	level_1	level_2	level_3	level_4
包含社区数量	1517	355	214	206	206

后续分析中,笔者选用了词频数大于 25 的 61 个高频关键词,图 1 列出了 level_4 的高频关键词聚类结果可视化效果图,将不同的社区标注了不同的颜色,以便区分。

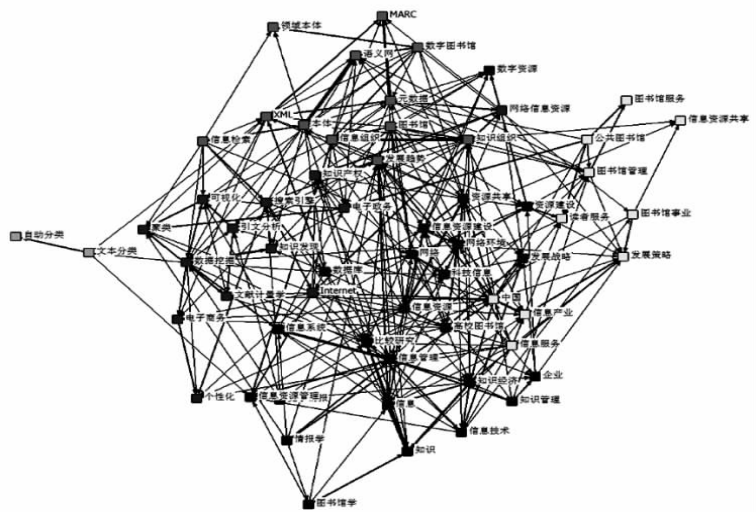


图 1 聚类结果可视化效果图

4.3 核心圈探测 按照 Z-value 的计算公式,对所划分的五个层级的每个社区中的节点进行 Z-value 分析,得出每个节点的 Z-value,并将每个社区中的关键词节点按照 Z-value 从高到低排序,将 Z-value 大于等于 2.5 的节点提取出来作为每个社区的核心圈。表 3 列出了 level_4 得到的三个社区 84、159 和 196 的 Z-value 情况,可以看到三个社区 Z-value 大于 2.5 的核心节点分别有 8、5、4 个,它们组成了每个社区的核心圈。

表 3 三个社区的 Z-value 情况

词汇	* Z-Value	词汇	* Z-Value	词汇	* Z-Value
图书馆	16.25132	知识管理	10.74367	数字图书馆	16.82191
信息服务	9.455315	信息	5.371834	知识产权	3.234983
信息资源	8.2734	知识	3.069619	网络信息资源	2.911485
网络环境	5.318615	知识服务	3.069619	互操作	2.587987
图书馆管理	4.727657	情报研究	3.069619	信息网络	1.94099
高校图书馆	3.841222	知识经济	1.918512	图书馆学教育	1.94099
资源共享	3.250264	知识共享	1.918512	信息组织	1.94099
网络	2.659307	知识网络	1.53481	著作权	1.94099
信息	2.06835	信息外延	1.53481	蛋白质数据库	1.617492
数学模型	2.06835	实验设计	1.53481	发展现状	1.617492
.....		

5 评价

5.1 聚类评价 笔者将从数量和质量两个方面来对关键词的聚类效果进行评价。

首先从数量上来评价。从表 2 中我们可以看到,用 Blondel 社区发现算法进行第一轮聚类之后,得到 level_0 的划分结果,即将 6 243 个关键词聚成了 1 517 个社区,平均每个社区含有 4.12 个关键词,这几乎相当于平均每篇文章的关键词数目,故我们可以得到,在第一轮聚类过程中,基本上将一篇文章的关键词聚成了一个社区,对少数文章进行了合并;从表中我们还可以看到,在进行第二次社区划分得到 level_1 的结果后,社区数量降到了 355,说明在这一轮中聚类中大部分社区合并为了一个社区,而在后面几轮的划分中,社区数量减少的很少,说明能够根据模块度升高而进行合并的社区已为数不多,到最后社区数量趋于稳定,即最终将 6 243 个关键词聚成了 206 社区,平均每个社区含有 30 个关键词。

在层次五(level4)上,6 243 个关键词聚成 206 个社区,显然在数量上过多,在分析了具体数据情况之后,笔者发现这 206 个社区里,仅仅只含有一篇或两篇文章的关键词的社区有 176 个,我们将这类社区称之为孤立社区,几乎没有关键词与这些社区的关键词相同,这可能是由于这些社区关键词使用不恰当或者所涉及学科内容比较小众造成的。孤立社区中包含 668 个关键词,占总关键词的 10.7%。除去孤立社区,剩余 30 个有意义的社区。某种意义上说,这 30 个社区

代表着图情领域的 30 个研究主题^[19]。

从 30 个社区中,笔者选取了三个社区 84、159、196 进行进一步具体分析,这三个社区包含的关键词数量分别 692、308 和 441,都是比较大的社区。图 2 是最终社区 84 的形成过程,图中的数字代表社区编号,可以看到,它由最初 level_0 的 121 个社区一层一层进行聚类,最终得到一个社区 84。

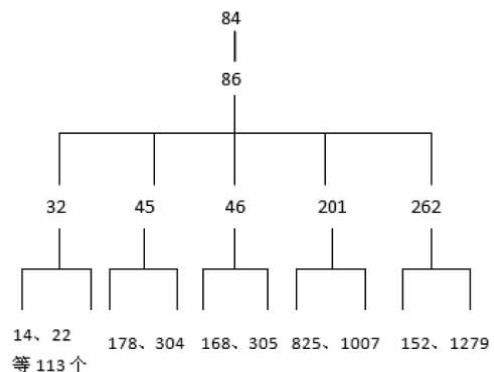


图 2 最终社区 84 的形成过程

从整体数量上的变化可以得到,Blondel 社区划分算法是有效的,它慢慢迭代着对整个网络进行合并,得到最终的五层结果。

从质量方面来进行评价。表 4 列出了三个社区的关键词列表。从(a)中可以看到,社区 84 的关键词集合代表的主要内容有图书馆、网络资源、情报信息等,将这些信息归为一类是可取的,都是信息资源方面的内容,但由表中可以看到,存在一些不能代表内容的关键词,如“伦理”、“中国西部地区”等,这些关键词的存在使得整个社区的内容变得更加难以分析;从(b)中可以看到,此社区主要描述的是知识管理、信息论方面的内容,归在一类也是很好的,但依旧存在一些不能归为此类的词汇,如“堵塞”、“治黄”、“河南”等;在(c)中,主要描述的研究内容为数字图书馆、网络信息资源等,同样可以归为一类,也存在诸如“标准”、“商标”等不合群的关键词。这些不合适的关键词的出现,有以下几方面的原因:一是文章作者在进行关键词标注时,使用不常用的关键词或者比较偏的表达方式;二是论文集中有一部分文章是属于跨两三个专业的,因此含有其他专业的关键词词汇;三是笔者在进行关键词规范化时,未能将所有无意义的关键词和同义异性词处理完全。

总的来说,从这三个社区划分的情况分别来看,Blondel 社区划分的算法是比较有效的,但从三个社区的整体来看,社区(a)84 和(c)196 的内容有相似之处,都与图书馆、信息资源相关,这两个社区的内容未能完全区分开来,可以进一步归类。

以上说明,从质量上来说,Blondel 社区划分算法

还存在不足之处,但总体来说效果非常好。

表 4 (a) 社区 84 的关键词列表

图书馆	信息服务	信息资源	网络环境	图书馆管理
高校图书馆	资源共享	网络	对策	数学模型
互联网	管理模式	印刷资源	发展策略	评审机制
信息资源共享	创新	人力资源	共建共享	开发利用
科技信息	管理机制	信息共享	伦理	档案馆
组织结构	信息需求	人事管理	可持续农业	措施
数字化	信息产业	资源建设	精神文明	体制改革
地方文献	共享模式	科技情报	人文精神	转型
信息资源配置	机制	科技文献	主动服务	开发
文献信息资源	图书馆功能	市场规划	电子资源	系统理论
文科核心期刊	优化	人的工作	信息服务业	文献分类
中国西部地区	优势	发展管理	在版编目	计算机
结构化综合布线	创新发展	用户	社会化	因特网站
战略组成要素	体系建设	度量	网络平台	域名保护
.....				

表 4 (b) 社区 159 的关键词列表

知识管理	信息	知识	知识服务	情报研究
知识经济	知识共享	展望	信息外延	实验设计
知识管理系统	风险管理	注意力	信息传递	知识创新
场依存-独立性	价值观	堵塞	信息经济	软件
Knowledge Works	网络技术	管理	XTM	分布式
主题文本资源	外部知识	技术基础	重要句群	流程重组
价值理论探索	信息变异	信息经济学	信息理论	建设目标
竞争情报软件	信息化水平	共享	研究现状	比较研究
知识生态环境	创新经济学	治黄	考核	知识转化
图书馆核心能力	相关度算法	知识价值	WebQL	产品责任
知识管理层次	可使用性	社科信息	方法基础	信息研究
广义 Ontology 方法	信息物理学	信息论	全球性整合	评价系统
TRS 赛迪数据	知本置换	河南	云块融合	知识对象
信息技术评价	可视化接口	组织知识	技术	响应
.....				

表 4 (c) 社区 196 的关键词列表

数字图书馆	知识产权	网络信息资源	互操作	信息网络
图书馆学教育	信息组织	著作权	DES 模型	发展现状
面向对象技术	多维索引	传统图书馆	互操作协议	资源管理
语义 Web 服务	长期保存	数字鸿沟	系统设计	资源整合
资源共建共享	国际图联	版权保护	标准	读者至上
系统基本模型	内容安全	中介读者	知识自由	保密法
图书馆社会化	电子图书馆	无线网络	生物信息学	研究现状
联合书目中心	法定呈缴	图像信息	章节组织	专业调整
蛋白质数据库	数字资源	管理与维护	虚拟图书馆	信息模型
图书馆自动化	特色资源	拓片数字化	图书馆立法	立场观点
电子商务服务	商标	开放描述	建设模式	多代理
技术接受模型	法定许可	法律关系	研制原理	对等网络
文献存取代理	培育模式	南京图书馆	可扩展性	中间件
中图法电子版	查询映射	光盘目录	中图法	样本赠送
.....				

5.2 核心圈评价 从上节的分析可知,社区 84 的主要内容为图书馆、网络资源、情报信息等,对社区 84 进行 Z - value 分析,得到 Z - value 大于 2.5 的关键词依次为“图书馆”、“信息服务”、“信息资源”、“网络环境”、“图书馆管理”、“高校图书馆”、“资源共享”和“网络”,可以看到,这些词汇描述的主要内容为图书馆、信息资源、网络等,与前面分析的整个社区内容大致相符,可以得到用这些 Z - value 大于 2.5 的词汇可

以描述整个社区 84 的结论;用同样的方法分析社区 159 和社区 196,得到社区 159 的 Z - value 大于 2.5 的核心词汇为“知识管理”、“信息”、“知识”、“知识服务”和“情报研究”,与前面分析的该社区描述的是“知识管理和信息论相关的内容”不谋而合;在社区 196 中,“数字图书馆”、“知识产权”“网络信息资源”和“互操作”这几个核心词汇描述的内容同样符合整个社区的主要内容。虽然这些社区的核心词汇无法描述出社区所有关键词的内容,但却可以覆盖大多数,故可以得到 Z - value 算法确定核心圈是行之有效的。

通过 Z - value 算法得到的结果,我们可以用“图书馆”、“信息服务”、“信息资源”、“网络环境”等词汇描述社区 84,可以用“知识管理”、“信息”、“知识”、“知识服务”等词汇描述社区 159,可以用“数字图书馆”、“知识产权”、“网络信息资源”和“互操作”来描述社区 196。

6 结 论

在对《中国图书馆学报》和《情报学报》两大期刊近 12 年的数据进行聚类 and 核心区探测之后,得到图书情报领域学科研究主题的具体聚类结果。笔者在对结果进行进一步提取之后,得到图书情报领域的学科研究主题可分为 15 个大类:信息计量、科学评价,竞争情报、战略分析,文献编目、期刊服务,数据挖掘、知识抽取,社会网络、知识地图,用户行为、协同标注,信息组织、知识管理,聚类分类、文献标引,数据挖掘、知识抽取,情报检索、信息查找,信息安全、信息法规,电子政务、政府信息,图书馆建设、数字图书馆,信息资源管理、数字信息资源,信息系统、计算机网络。这 15 个大类涵盖了图书情报领域的大部分主题,这些主题之间存在一定的交叉性,是因为学科主题都是相关联的,加上所用的方法也存在一定局限性。

在本文中,笔者运用共词分析的方法来构建网络,与引文网络、合著网络等有着本质的区别,在此基础上,笔者将社会网络的分析方法合理运用到聚类算法中,将 Blondel 社区划分算法与 Z - value 社区核心圈确定的方法有效结合起来,对整个学术文档的关键词网络社区进行分析,从而完成学科主题的聚类,最终得到了图书情报领域的 15 大类别,并对聚类算法和核心圈发现算法进行了评价。

本文在研究过程中存在一些不足,在今后的研究中,笔者将对关键词进行更加规范的处理,并引进其他的聚类算法对构建的网络进行主题聚类分析,以比较不同聚类算法的优劣,在学科数据的选择上,也将选择不同学科领域更多的数据进行分析,进一步验证这些算法的准确性。

(下转第 39 页)

系较弱,这使得核心技术领域的研究更具有现实意义。研究结果证明,“m-核”分析方法能够有效地识别共现网络中依赖程度最大、共性关系最强的核心技术领域。本文所提出研究方法对于核心技术的识别也颇有裨益。掌握产业或技术主题的核心技术领域、核心技术,能够为相关单位技术研发策略提供决策支持,对于提高国家(地区)的国际竞争力、推动技术和技术发展发挥着巨大作用。

我们下一步的研究方向是将时间序列加入共现网络中,通过分析技术应用领域的动态共现网络图,挖掘和还原技术领域研发动态和技术领域的变迁过程,明确技术发展趋势;其次,对核心技术领域进行技术创新信息的挖掘,从而确定技术领域所需要突破的技术壁垒。

参 考 文 献

- [1] 栾春娟,曾国屏. 基于 SNA 核心技术领域测度研究[J]. 图书情报工作, 2011, 55(6): 33-52
- [2] Liu C, Yang J. Decoding Patent Information Using Patent Maps[J]. Data Science Journal, 2008, 7: 14-22
- [3] 张存刚,李明,陆德梅. 社会网络分析——一种重要的社会学研究方法[J]. 甘肃社会科学, 2004(2): 109-111
- [4] 沈君,王续琨,高继平,等. 基于文献计量指标的关键技术的探寻——以第三代移动通信技术为例[J]. 情报杂志, 2011, 30(9): 34-40
- [5] Kwon O, Seo J, Noh K, Kim J, Kim JS, et al. Categorizing Influential Patents Using Bibliometric Analysis of Patent Citations Network[J]. Information, 2007, 10(3): 313-326
- [6] Lee Y G, Song Y I. Selecting the Key Research Areas in Nano-technology Field Using Technology Cluster Analysis: A Case Study Based on National R&D Programs in South Korea[J]. Technovation, 2007, 27(1-2): 57-64
- [7] 俞文华. 美国在华技术比较优势演变及其政策含义[J]. 科学学研究, 2008, 26(2): 98-104
- [8] 林聚任. 社会网络分析: 理论、方法与应用[M]. 北京: 北京师范大学出版社, 2009: 134-137
- [9] 谢彩霞,梁立明,王文辉. 我国纳米科技论文关键词共现分析[J]. 情报杂志, 2005, 24(3): 69-73
- [10] 沈君,高继平,滕立. 德温特手工代码共现法: 一种实用的专利地图法[J]. 科学学与科学技术管理, 2012, 33(1): 12-16
- [11] Penan H. R&D Strategy in a Techno-Economic Network: Alzheimer's Disease Therapeutic Strategies[J]. Research Policy, 1996(25): 337-358
- [12] 唐晨辉. 变频技术及其应用[J]. 江西科技师范学院学报, 2002, 6(3): 57-59
- [13] 刘向阳,刘杰英. 变频技术的普及前景及其技术保证[J]. 信息技术, 2002(2): 20-22
- [14] 王智. 变频技术在家用空调中的应用[J]. 黑龙江科技信息, 2008(5): 28-29

(责编: 王平军)

(上接第 44 页)

参 考 文 献

- [1] Newman M E J, Girvan M. Finding and Evaluating Community Structure in Networks[J]. Phys Rev E, 2004, 69: 26-113
- [2] Latapy M, Pons P. Computing Communities in Large Networks Using Random Walks[C]. //Proc of the 20th International Symposium on Computer and Information Science, 2005: 284-293
- [3] Palla G, Derenyi I, Farkas I. Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society[J]. Nature, 2005, 435: 814-818
- [4] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast Unfolding of Communities in Large Networks[J]. Jstat Mech-theory, 2008
- [5] Rosvall M, Bergstrom CT (2008) Maps of Information Flow reveal Community Structure in Complex Networks[J]. PNAS, 105: 1118-1123
- [6] Martin Rosvall, Carl T Bergstrom. Mapping Change in Large Networks[J]. PLoS One, 2010, 5(1). arXiv: 0812. 1242v1
- [7] 林颂坚. 基于术语抽取与术语群技术的主题抽取[J]. Computational Linguistics and Chinese Language Processing, 2004, 9(1): 97-112
- [8] 张晗,崔雷. 运用共词聚类分析研究生物信息学的研究热点[J]. 医学情报工作, 2004, 25(5): 327-330
- [9] 黄小燕. 情报领域研究热点透视——情报领域论文关键词词频分析(1999-2003)[J]. 图书与情报, 2005(6): 82-84
- [10] 马费成,张勤. 国内外知识管理研究热点——基于词频的统计分析[J]. 情报学报, 2006(2): 163-171
- [11] 屈鹏,隆捷,吴龙婷,等. 国内外情报学研究现状的统计分析[J]. 情报资料工作, 2007(3): 82-86
- [12] 王翼,杜楠,吴斌. 复杂网络在文献信息服务中的应用于实现方法[J]. 数字图书馆论坛, 2008(6): 34-37
- [13] 章成志,梁勇. 基于主题聚类的学科研究热点及其趋势监测方法[J]. 情报学报, 2010, 29(2): 342-349
- [14] Costa L F, Rodrigues F A, Traviesso G, et al. Characterization of Complex Networks: A Survey of Measurements[J]. Advances in Physics, 2007, 56(1): 167-242
- [15] Newman M E J. Fast Algorithm for Detecting Community Structure in Networks[J]. Phys. Rev. E, 2004, 69: 066133
- [16] Mislove A, Marcon M, Gummadi K P, et al. Measurement and Analysis of Online Social Networks[C]. Proc of the 7th ACM SIGCOMM Conference, 2007: 29-42
- [17] Guimerà R, Sales-Pardo M, Amaral L A N. Classes of Complex Networks Defined by Role-to-role Connectivity Profiles[J]. Nature Physics, 2007(3): 63-69
- [18] 冯文敏. 学术论文摘要与关键词规范化浅析[J]. 武警学院学报, 2011, 27(3): 93-96
- [19] 王晓光. 科学知识网络的形成与演化(I): 共词网络方法的提出[J]. 情报学报, 2009, 28(4): 599-605

(责编: 贺小利)