

doi:10.3772/j.issn.2095-915x.2015.04.009

基于专利 IPC 分类号的技术竞争对象的群组分析方法

韩红旗, 付媛, 朱礼军

(中国科学技术信息研究所 北京 100038)

摘要: 技术竞争对象群组是产业技术内的技术研发实体因为技术上接近而形成的集群。发现技术竞争对象群组, 可以从较高层次上掌握产业技术领域内的竞争态势, 给产业技术参与者和政策制定者提供有价值的情报信息。首先概述了主要的群组分析方法, 然后引入了可视化的 LinLog 图形聚类分析方法, 从研发机构、省份和国家三个层面, 提出了以 IPC 分类号为特征的技术竞争对象群组的识别流程模型。采集了燃料电池技术领域的专利数据作为实验研究, 对数据进行了预处理, 采取提出的方法对该领域的竞争对象群组进行了识别, 验证了该方法的有效性。

关键词: IPC 分类号, 竞争对象群组, LinLog 算法, 专利分析

中图分类号: G35

Patent Technology Competitor Group Analysis Method Based on IPC

HAN Hongqi, FU Yuan, ZHU Lijun

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

Abstract: It is crucial to understand the technical groups of intra-industry and to master the competition in the field of technology. In order to provide valuable information to industry participants and policymakers, the

基金项目: 本研究得到国家自然科学基金资助项目(编号:71473237)和“十二五”国家科技支撑计划课题“专利信息资源挖掘与发现关键技术研究”(2013BAH21B02)资助。

作者简介: 韩红旗(1971-), 博士, 副研究员, 研究方向: 文本挖掘、社会网络分析、科技情报分析, bithhq@163.com; 付媛(1989-), 硕士, 中国科学技术信息研究所, 研究方向: 知识工程, 数据分析; 朱礼军(1973-), 博士, 研究员, 研究方向: 语义网、网络服务和知识技术在科技信息服务、电子政务/商务中的应用、知识组织系统。

Patent technology competitor group analysis method based on IPC is introduced. Firstly, the analysis method of groups is summarized. The recognition model of technological competitor groups characterized by the IPC classification number and the technical terms is put forward. Then, obtain patent data for this research in the field of fuel cell technology through multi-threaded Chinese patent data acquisition system. Finally, invoking LinLog visualization tools to do graphical clustering for research and development institutions, provinces and countries and the patent technology competitor analysis method is validated.

Keywords: IPC classification number, technology competitor group, LinLog algorithm, patent analysis

1 引言

一般来说, 产业技术是竞争性技术^[1], 在一段时间内, 总有一些技术受到追捧, 成为产业内的优势技术或主导技术。集聚是产业的一种常见形态, 当产业技术集聚到一定程度、能够较好地满足产品功能需要时, 产业的发展就进入成熟阶段, 集聚起来的各种技术以一定的结构形态形成了产业技术体系。另一方面, 产业技术的存在有赖于拥有产业技术的对象, 它们相互竞争形成了技术上接近的群组。按照波特的竞争优势理论, 在产业内, 真正的竞争对手是与自己类似的厂商, 这些厂商构成一个战略群, 成为次产业, 产业内不同的战略群之间存在移动壁垒, 因此具有相似产业技术的企业很可能是产业内最主要的竞争对手^[2]。

将分析的数据划分为数个簇的聚类方法能使用户发现数据中的关系模式, 了解隐藏在数据中的知识。产业技术的竞争对象群组分析就是采用适当的聚类方法, 通过对数据的分析, 将竞争对象划分为若干个组, 从而识别产业内相互竞争、相互影响的竞争对象集群的一种方法^[3]。技术竞争对象群组分析可以从国家、省份和研发机构等不同层面了解产业内的技术群, 从较高层次上掌握产业技术领域内的竞争态势, 给产业技术参与

者和政策制定者提供有价值的情报信息^[4]。

识别产业技术中的竞争对象群组可以采用一些常用的聚类技术^[5], 如自组织映射、K-means、因子分析技术等^[6]。具体做法是将每一个技术竞争对象表示为若干技术特征度量的特征向量, 通过计算对象之间的距离来将相近的对象聚为一类。文献研究发现^[7], 更多的研究者采用了可视化分析方法。传统聚类算法因为无监督学习的特点, 常使人们怀疑其分析结果的有效性, 而可视化方法由于有效地结合了计算机技术和人类认知能力, 将抽象的数据用图形形式展示出来, 给用户以直观、形象的感觉, 增强了用户对分析结果的信心, 近些年受到了广泛推崇。

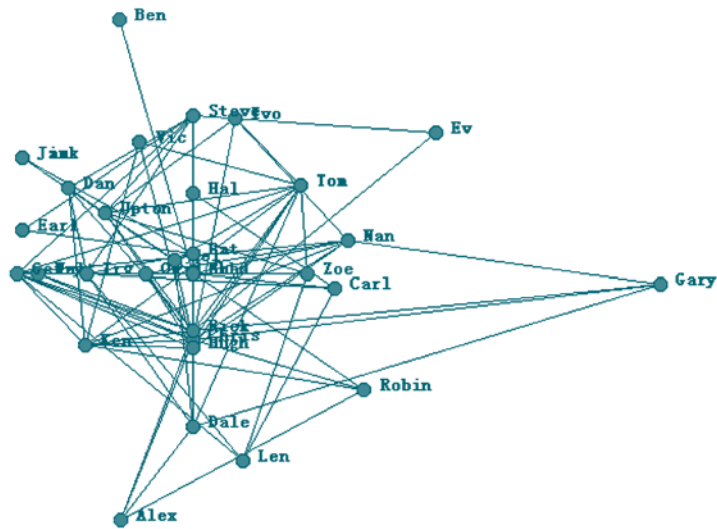
2 相关知识

2.1 LinLog 图形聚类方法

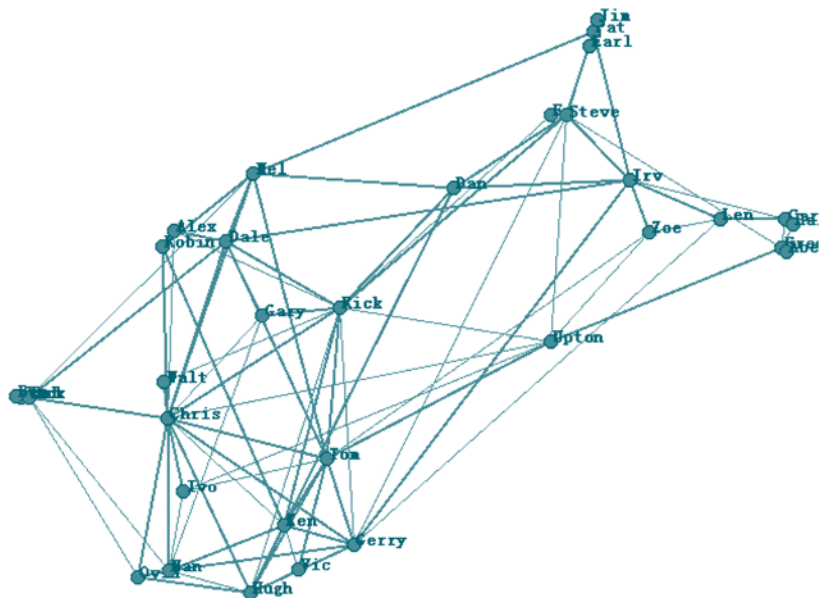
鉴于可视化方法具有直观、易于理解的优势, 本文采用可视化方式来发现不同层次下的技术竞争对象群组。然而, 现在广为采用的能量模型只是产生易理解的可视化图 (readable visualization), 而不是为聚类进行设计的。这些模型强调较少的和均衡的边长, 目标在于绘制出漂亮美观的图形, 所以它们倾向于将度数大的结

点放在布局图的中央,因为这能够使它们距离其它相应结点的距离最短,而这常常阻止将分离的结点分到不同的簇中,因为分开簇需要较长的边。显然,如果根据传统绘图算法表示的图形来是识别技术竞争对象群组,可能会产生错误的结果,因此本文采用 LinLog 图形聚类算法而不是 Spring 算法来识别技术竞争对象群组。LinLog 方法是德国勃兰登堡科技大学的 Noack^[8-10] 提出的,其目

标在于产生理想的可视化聚类效果图形。图 1 是 Noack 论文中提到的一个例子,分别采用 Spring 方法和 LinLog 方法对其数据进行图形绘制。从绘制的图形来看,可以看到 LinLog 方法较清楚地将图形分为两个大的簇,两个簇间通过 Dan 和 Upton 两个结点连接,而 Spring 方法则将度数大的结点放在布局图中央,将度数小的结点放在布局图外围,没有将数据很清楚地分开。



(a) Spring 模型



(b) LinLog 模型

图 1 Spring 方法和 LinLog 方法绘制图形的对比

LinLog 能量模型方法不符合传统的美学标准，它的目标是布局时将紧密连接的结点分组，而将稀疏连接的结点分离。LinLog 模型有结点斥力模型 (Node-RepulsionLinLog) 和边斥力模型 (Edge-RepulsionLinLog) 两个算法，分别基于两种著名的聚类标准，即割密度 (Densityofthecut) 和规范割 (NormalizedCut)。NormalizedCut 和 Edge-RepulsiveLinLog 在具有较高度的结点分组时能产生无偏的结果，因此特别适合于实际常见的右斜分布度的图。Noack 的算法在提出后进行了多次改进，本文绘制的聚类图采用的 LinLog 算法是 Barnes 和 Hut 层级算法，算法执行完后不仅可以绘制图形，而且可以将结点划分为若干个簇。为了更清晰地展现聚类结果，在绘制图形时采用不同的结点颜色来区分簇的方法，即同一簇的结点采用同一颜色，而不同簇的结点颜色各不相同。

2.2 IPC 分类系统简介

国际专利技术分类号 (International Patent Classification, IPC) 是专利文献扉页中专利审查员赋予专利的技术分类号，一个专利至少有一个 IPC 号，但并不限于一个 IPC 分类号，也就是说有些专利被赋予两个以上的 IPC 分类号，当专利存在多个分类号时，第一个分类号称为主分类号。IPC 是目前惟一国际通用的专利文献分类和检索号，是世界各国专利局都采用的国际标准，虽然一些国家或组织制订了自身的专利分类系统，如美国专利商标局 (USPTO) 制定的 UPC 专利分类号，欧洲专利局 (EPO) 制定的 ECLA 分类号，但它们同时也提供了 IPC 分类号，我国专利也采用了 IPC 技术分类号体系。

IPC 是按照发明创造的技术主题为特征进行分类的^[11]，将不同的技术领域概括分成 8 个部，每个部代表不同技术类别，分别用英文字母 A~H 表示。各部对应的技术领域见 (表 1)。

表 1 部与技术类别

部	对应的技术类别
A	人类生活必需
B	作业，运输
C	化学，冶金
D	纺织，造纸
E	固定建筑物
F	机械工程，照明，加热，武器，爆破
G	物理
H	电学

资料来源：陈燕等^[12]；李建蓉^[13]

IPC 的分类体系是由高到低依次排列的等级式结构，把与发明创造有关的全部技术领域按不同的技术范围设置成部、大类、小类、大组或小组，由大到小的递降次序排列。每一个部分为若干类，用表示部的字母加上两个阿拉伯数字表示，如 A01，所表示的技术类别也进一步细化。类下面是小类，相应的表示形式是在类的后面加上一个英文字母，如 A01B。进一步，小类又分为组，



图 2 IPC 命名规则与类别描述示意图

资料来源：刘玉琴^[14]

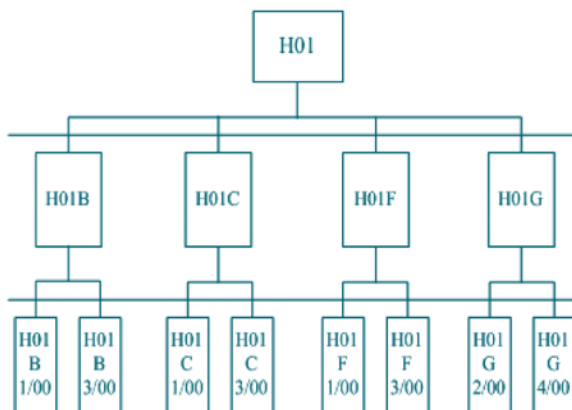


图 3 IPC 层次结构示意图

资料来源：刘玉琴^[14]

包括大组(主组)和小组(子组),如(图2)所示,从而构成了一个层次结构,该结构以部为根节点,以子组为叶结点,当到组的这个层次时已经有近70000个技术类别,所以这个系统是非常庞大的,(图3)是该层次结构的一个片段。

3 基于 IPC 的竞争对象群组识别流程模型

一个产业技术领域一般可以划分为若干个子领域,每个子领域往往又存在很多更小的技术子领域,不同的技术竞争对象有着各自不同的技术背景和基础,研发目标和重点各不相同,在同一子领域下进行技术研发的技术竞争对象,既可能是市场上的竞争者,又可能是合作者,也是相似技术领域的相互影响者。专利审查员为专利标引的 IPC 分类号,为了解产业技术热点和不同技术竞争对象的研发方向提供了有效的途径^[15]。然而,一个技术竞争对象往往在多个技术子领域开展研究,单从 IPC 计数统计很难判断两个技术竞争对象的研发技术是否接近。因此提出一种以 IPC 主分类号作为特征的图形聚类方法,通过可视化形式识别产业技术领域内的技术竞争对象群组,图4是此方法的流程模型。

首先选择聚类层次类别,可以选择研发机构、省份和国家三种。然后统计选定层次下每个技术竞争对象在每个 IPC 分类号下的专利数量,从而可以建立技术竞争对象与主 IPC 分类号之间的关系矩阵,将每个技术竞争对象表示为以主 IPC 分类号作为特征的特征向量,向量的每一个特征值用该对象在此主 IPC 分类号下的专利数量表示,接下来采用定义的特征向量计算不同技术竞争对象间的相似度。也有研究人员采用相关系数作为可视化图形绘制的输入数据,如杨璧嘉^[16],这里采用相似度而没有采用相关系数,是因为相关系

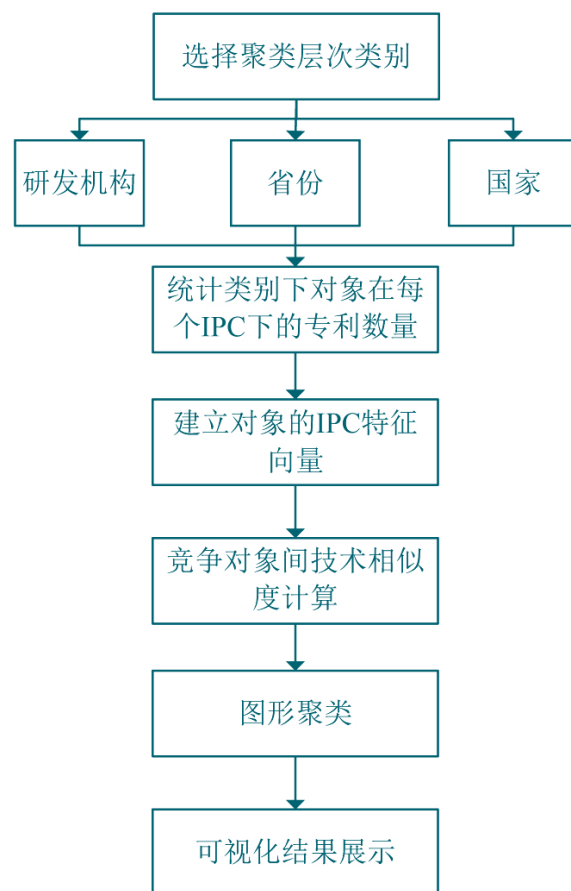


图4 基于 IPC 的技术竞争对象图形聚类方法流程模型

数可能存在负值,采用的图形聚类算法不支持边的权重为负值的情况。设产业技术涵盖的 IPC 主分类号列表数量为 $|IPC|$, 对象 i 对应的第 k 个 IPC 分类号的特征值为 IPC_{ki} , 即对象 i 在第 k 个 IPC 分类号下的专利数量, 则两个技术竞争对象 i 和 j 的技术相似度的计算见公式 1:

$$sim(i, j) = \frac{\sum_{k=1}^{|IPC|} IPC_{ki} \times IPC_{kj}}{\sqrt{\sum_{k=1}^{|IPC|} IPC_{ki}^2} \times \sqrt{\sum_{k=1}^{|IPC|} IPC_{kj}^2}} \quad (1)$$

为了获取较好的可视化图形,一般需要设定一个最小的相似度阈值,阈值一般取相似度的均值,也可通过实验确定。只有当两个技术竞争对象的技术相似度高于此阈值,才认为它们之间存在连接边^[17]。以技术竞争对象为结点,它们之间

的连接为边，边的权重取它们之间的技术相似值，调用 LinLog 图形聚类算法，即可取得以可视化图形方式展示的技术集聚图，并识别出不同层次下的技术竞争对象群组。

4 基于专利 IPC 分类号方法的竞争对象的群组分析

4.1 数据的获取和预处理

现在，世界上一些主要国家和国际性组织的专利局在自己的官方网站上提供了可免费检索的专利数据库和检索程序，人们通过互联网可以随时地连接到专利局的网站，实时地获取需要的专利数据。为了验证提出的技术竞争分析方法，利用专利采集系统（图 5）从中国国家知识产权局官方网站下载了燃料电池技术领域的专利数据，共下载专利 6346 件。获取的数据为包含专利著录项目信息的 HTML 格式的网页，通过网页内容分析后将专利著录项目数据抽取出来并存入本地数据库。

采集到的专利数据往往存在一些问题，在进行正式分析前需要经过一个数据预处理过程才能

较方便地使用。下面对本研究涉及的主要专利数据预处理方法进行介绍，主要包括识别专利的类别，识别申请（专利权）人的国家、省份，识别申请（专利权）人的类别，对不同名称但实属一个组织的专利进行合并等。

4.1.1 专利类别的识别

中国专利的类别分为发明专利、实用新型专利和外观设计专利三种。既有本国专利权人申请的国内专利，也有通过中国国家知识产权局进行 PCT 申请的专利，也有在其它国家进行 PCT 申请而进入中国国家阶段申请的专利。

中国专利的申请号由 8-13 位数字组成，专利申请号中包含了专利所述的类别。中国各种专利说明书自 1985 年 9 月 10 日开始出版以来，其编号体系在 1989 年、1994 年、1998 年和 2003 年先后进行了四次调整，相应地可以把专利申请号的使用情况分为五个阶段，详细情况见表 6.1。第一阶段从 1985 到 1989 年，申请号为 8 位长度，第 3 位表示了专利的类别；第二阶段从 1989 年到 1993 年，申请号为 9 位长度，只是在原来长度后面增加了计算机校验位；第三阶段从 1994 年到 1997 年，后 5 位申请号以 8 和 9 开头的表示 PCT 国际申请；第四阶段从 1998 年到 2003 年 9 月，

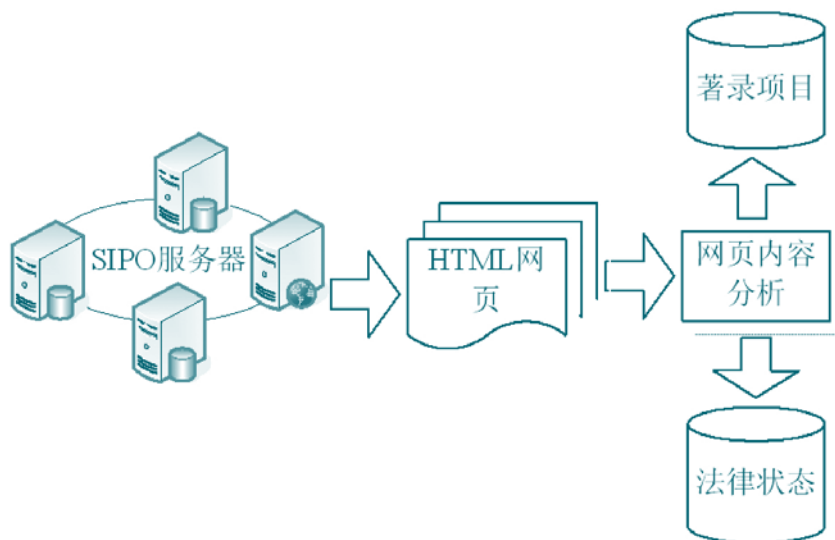


表 2 中国专利申请号格式

使用阶段	申请号长度	特点	示例
1985-1988	8 位	前两位为年代；第 3 位为专利种类，1 代表发明专利，2 代表实用新型，3 代表外观设计	85100244
1989-1993	9 位	开始在申请号增加了小数点后面的计算机校验位。	90103974.8
1994-1997	9 位	发明专利申请号后 5 位申请序号中，以 8 和 9 开头的，表示指定中国的国际申请	96180120.4
			97192636.0
1998-2003.9	9 位	申请号中第 3 位为 8 或 9 的表示 PCT 国际申请	98801694.X
2003.10- 现在	13 位	前 4 位为申请年代，第 5 位为专利申请类型，第 6-12 位为当年的顺序号	200410002053.3
			200480001636.2

申请号中第 3 位为 8 和 9 的表示 PCT 国际申请；第五阶段从 2003 年 10 月到现在，申请号前 4 位为年份，第 5 位代表了专利的类别。根据（表 2）所提供的信息可以确定每一件专利的类别。

4.1.2 申请（专利权）人国家和省份的识别

对于第一申请（专利权）人为中国的专利，著录项目中的地址项包含了其所属的省份信息。地址著录项目的前 6 位数据是申请（专利权）人的邮政编码，可以根据邮政编码查询邮政编码数据库识别出省份信息。对个别地址项目中没有邮政编码或在邮政编码数据库中检索不到的专利数据，则可根据地址数据中包含的省、市等信息来识别。

对于第一申请（专利权）人为国外个人或组织的在中国申请的专利，专利著录项目中的优先权项、国际公布项包含了这些国外个人或组织所属的国家信息。例如一个专利的优先权项是“1999.8.27 JP 242132/1999”，其中的 JP 表示了这是一个日本国家的申请（专利权）人；又如另一件专利不存在优先权项，但国际公布项是“2003-12-12PCT/KR2003/002728”，则其中的 KR 表示此专利权人属于韩国。两个例子的优先权项或国际公布项中的双字母代码代表了国家、其它实体或政府间组织。本研究所涉及的国家或组织双字母代码及对应的名称见附录 B，完整的国家、其它实体或政府组织的双字母代码参见 <http://www.sipo.gov.cn/>。

4.1.3 申请（专利权）人类别的识别

为了研究上的目的，将申请（专利权）人分为公司、大学、研究所、其它和个人 5 类，申请（专利权）人所属类别主要以名称中包含的关键词来确定。申请（专利权）人名称中出现的关键词和类别的对应关系见（表 3）。当一个专利有多个申请（专利权）人时，以第一个申请（专利权）人的类别作为此专利的统计类别，如“00112136.7”号专利的申请（专利权）人有两个，分别为：南京师范大学和中国科学院长春应用化学研究所，则以第一个申请（专利权）人的类别即“大学”作为该专利的申请（专利权）人类别。在一个名称中出现两个类别的关键词时，则取名称中靠后位置的关键词为依据。在通过关键词确定申请（专利权）人的类别后，还要采用一些手工手段进行检查和修正错误的分类。

4.1.4 异名同指处理

异名同指问题指的是申请（专利权）人虽然名称不同，但确实是一个组织或同属于一个大的跨国企业。造成这种状况的原因是多方面的，有些是同一家企业两次采用了相似但不同的名称，有的是名称翻译或标注时造成的，如“巴斯夫”和“巴斯福”，“国际燃料电池公司”、“国际燃料电池有限公司”和“国际燃料电池有限责任公司”；还有些是虽属于一个大的跨国企业，但分属不同的子公司，如株式会社日立制作所，日

立化成工业株式会社，日立自动车株式会社，日立麦克赛尔株式会社，日立家用电器公司等，虽然名称各不相同，但它们都是日立旗下的企业，因此在分析中要把它们归并为一个组织处理。

表3 申请(专利权)人类别识别关键词

类别	关键词
公司	公司, 合伙, 株式会社
大学	大学, 学院
研究所	研究所, 研究会, 研究院, 研究总院, 研究中心
其它	委员会, 中心, 协会, 基金, 部队
个人	

4.2 实验结果和讨论

分别选择了研发机构、省份和国家三个层次。首先统计选定层次下的每个技术竞争对象在每个IPC分类号下的有效授权专利数量，建立以IPC分类号为特征向量的矩阵，采用相似度计算公式计算每对对象的技术相似度，选择相似度均值作为阈值，舍弃低于阈值的数据，调用开发的LinLog可视化工具进行了图形聚类，分析结果如下：

4.2.1 研发机构

选择了专利有效授权数量排名前20的研发机构(第一专利权人)进行了图形聚类分析，分析结果见(图6)，图中结点的大小表示了有效授权专利数量，结点的颜色表示了其所属于的群组。上海神力科技有限公司的有效授权专利总数量为194件，对应了最大的结点，UTC公司的有效授权专利总数量为19件，对应了图中最小的结点。

共识别出两个技术竞争对象群组。结点颜色为红色的对应了第一群组，包括10个研发机构，分别是：三星公司(177)、中国科学院(128)、胜光科技股份有限公司(74)、通用汽车(56)、本田技研工业株式会社(52)、武汉理工大学(49)、上海交通大学(38)、三洋公司(37)、比亚迪股份有限公司(32)和哈尔滨工业大学(26)；结点颜色为橙色的对应了第二群组，包括另外10个研发机构，分别是：上海神力科技有限公司(194)、松下公司(154)、丰田公司(120)、清华大学(72)、日产公司(62)、东芝公司(48)、新源动力股份有限公司(26)、日立公司(24)、乐金电子(天津)电器有限公司(20)和UTC公司(19)。括号中的数字是各专利权人的

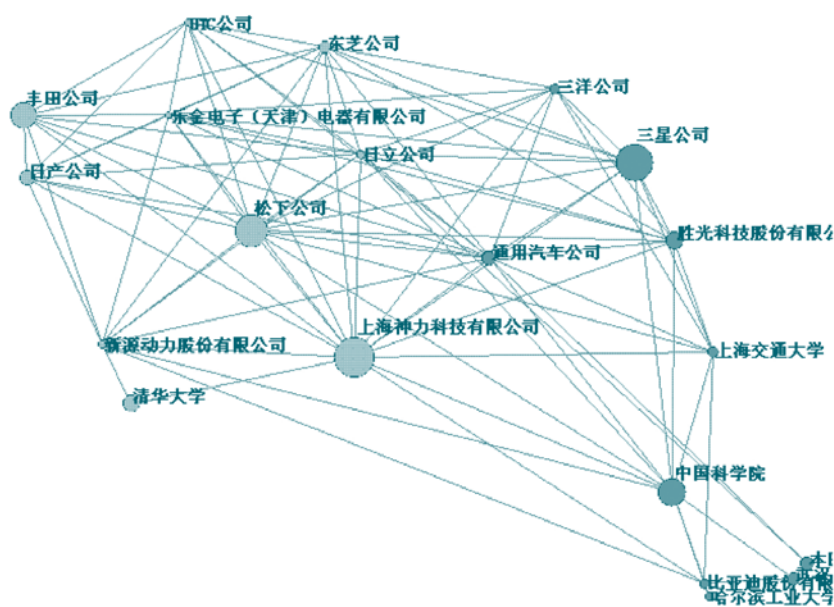


图6 IPC方法的研发机构聚类图

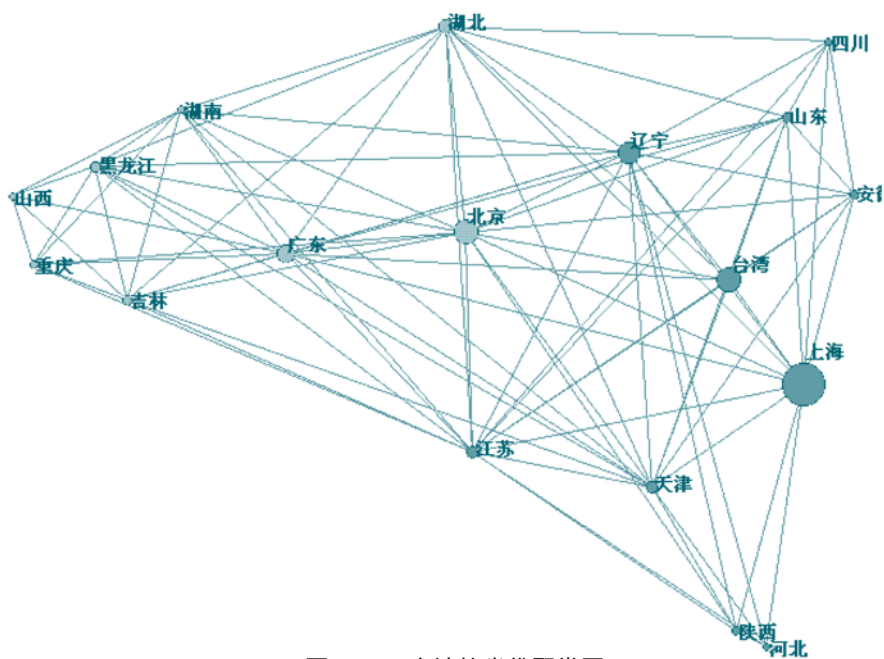


图 7 IPC 方法的省份聚类图

有效授权专利总数量。

4.2.2 省份

选择全部共 22 个拥有有效授权燃料电池专利的省份作为分析对象，分析结果见（图 7）。图中最大节点是上海，对应的有效授权专利总数量是 311 件，最小的节点是河北，对应的有效授权专利总数量是 2 件。

共识别出两个群组，第一群组用红色结点表

示，包括上海 (311)、台湾 (152)、辽宁 (127)、江苏 (41)、天津 (40)、山东 (23)、安徽 (19)、陕西 (13)、四川 (4) 和河北 (2) 等 10 个省份；第二群组用橙色结点表示，包括北京 (150)、广东 (93)、湖北 (58)、黑龙江 (29)、吉林 (18)、重庆 (5)、湖南 (4) 和山西 (4) 等 8 个省份。而浙江 (16)、福建 (12)、内蒙古 (1) 和云南 (1) 四个省份由于与其它省份的技术相似度低于阈值而没有在聚类图上表示出来。

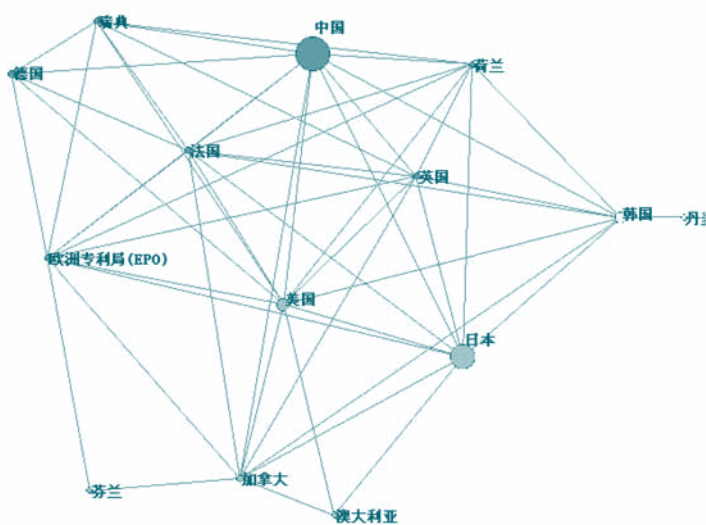


图 8 IPC 方法的国家聚类图

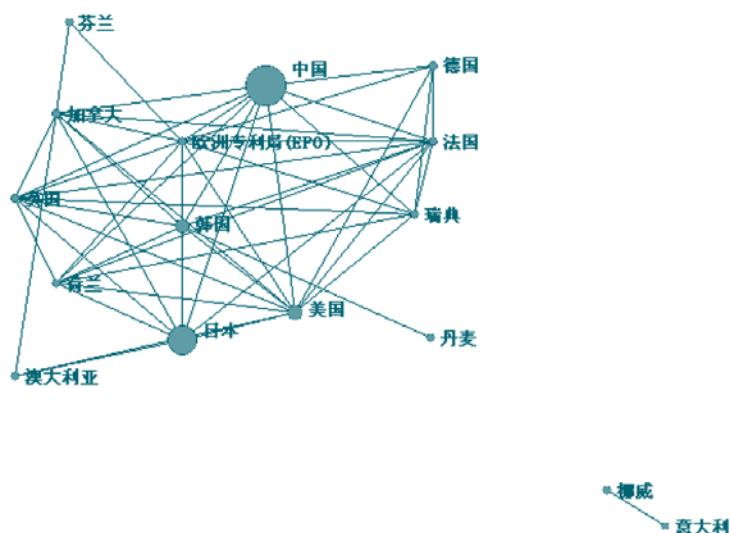


图9 非连通的国家对象关联图形

括号中的数字是各省份的有效授权专利总数量。

4.2.3 国家

选取全部共 17 个拥有有效授权专利的国家或地区性组织作为分析对象进行了技术相似性图形聚类, 分析结果见 (图 8)。图中最大结点是中国, 有效授权专利数量是 1123 件, 是有效授权专利数量最多的国家; 最小的结点是丹麦和芬兰, 对应的有效授权专利数量是 3。

共识别出四个技术竞争对象群组, 包含了 16 个国家或地区性组织。红色结点代表第一群组, 包括中国 (1123)、德国 (58)、英国 (28)、法国 (16)、欧洲专利局 (EPO)(10)、瑞典 (6) 和荷兰 (5) 等 7 个国家或地区性组织; 橙色结点代表第二群组, 包括日本 (740)、美国 (292)、加拿大 (33)、澳大利亚 (4) 和芬兰 (3) 等 5 个国家或地区性组织; 第三群组包含韩国 (202) 和丹麦 (3) 两个国家; 第四群组包括挪威 (3) 和意大利 (1) 两个国家。由于挪威和意大利两个结点对象之间存在边的连接, 但与其它结点对象之间不存在边 (图 9), 而目前的 LinLog 可视化算法在对非连通图生成图形时存在问题, 所以在生成图 6.32 聚类图形时做了特殊处理。奥地利 (1) 因为与其它国家或地区性组织的技术相似度低于阈值, 而不包含在聚类图中。

5 结束语

本文基于专利 IPC 分类号方法的竞争对象的群组分析, 采用了燃料电池技术领域的专利数据, 首先统计选定层次下的每个技术竞争对象在每个 IPC 分类号下的有效授权专利数量, 建立以 IPC 分类号为特征向量的矩阵, 采用相似度计算公式计算每对对象的技术相似度, 选择相似度均值作为阈值, 舍弃低于阈值的数据, 调用开发的 LinLog 可视化工具, 对研发机构、省份、国家三个层次进行了图形聚类。

参考文献

- [1] 郭丽. 企业技术战略多维选择模型研究 [D]. 同济大学, 2008.
- [2] 陈立新, 梁立明. 技术领域的集成与整合研究 [J]. 情报杂志, 2015, 32(1): 37-41.
- [3] 金泳锋. 中国矿用风机产业技术竞争态势研究 [J]. 情报杂志, 2013, 34(1): 50-54.
- [4] 李纪珍, 刘摇璇, 文庭孝. 国内专利竞争情报研究综述 [J]. 情报探索, 2013(11):57-63.
- [5] 任智军, 朱东华, 荆雷. 基于可视化数据挖掘的管理科学科技文本分析研究 [J]. 科学学与科学技术管理.

2006(1): 8-12.

[6] 赖玉霞, 刘建平. K_means 算法的初始聚类中心的优化 [J]. 计算机工程与应用, 2008, 44(10): 147-149.

[7] 王鑫, 赵蕴华, 高芳. 基于分类号和引文的专利相似度测量方法研究 [J]. 数字图书馆论坛, 2015, 128(1): 57-61.

[8] NOACK A. Energy models for graph clustering[J]. Journal of Graph Algorithms and Applications, 2007, 11(2): 453-480.

[9] NOACK A. Energy-based clustering[C]//13th International Symposium on Graph Drawing, 2005.

[10] NOACK A. An energy model for visual graph clustering[C]// 11th International Symposium on Graph Drawing, 2004.

[11] 鲍志彦, 张红芹. 基于专利地图的竞争情报挖掘及实证研究情报杂志 [J]. 图书情报工作, 2011(9): 20-23.

[12] 陈燕, 黄迎燕, 方建国. 专利信息采集与分析 [M]. 北京: 清华大学出版社, 2006.

[13] 李建蓉. 专利信息与利用 [M]. 北京: 知识产权出版社, 2006.

[14] 刘玉琴. 基于专利检索与专利分析的技术创新管理方法研究 [D]. 北京理工大学, 2008.

[15] Lee Sungjoo, YoonByungun, ParkYongtae. An approach to discovering new technology opportunities: Keyword-based patent map approach[J]. Technovation. 2009, 29(6-7): 481-497.

[16] 杨璧嘉, 张旭. 专利网络分析在技术路线图中的应用 [J]. 情报分析与研究, 2008(5): 61-66.

[17] YOON B, PHAAL R, PROBERT D. Structuring technological information for technology roadmapping: data mining approach[C]// 7th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED'08), 2008.