

布氏定律在燃料电池能量管理技术 专利信息离散分布中的应用探讨

苑朋彬 赵蕴华 张维冲

(中国科学技术信息研究所, 北京 100038)

摘要: 揭示和研究信息离散分布现象是情报学的重要课题。文章以燃料电池能量管理领域的相关专利数据为基础, 通过理论法、区域法和图像法, 探讨将布氏定律研究思路运用到专利信息离散分布现象的可行性问题, 总结出专利信息分布的特点。研究发现, 专利信息同期刊信息一样存在明显的集中与分散现象, 但集中和分散程度更强, 在一定程度上可以将布氏定律的研究思路应用到专利信息的离散分布现象研究中。

关键词: 布氏定律; 专利; 离散分布; 区域法; 图像法; 斯马利科夫统一方程

中图分类号: G350

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2018.02.009

Bradford Law Application for Patent Information Discrete Distribution in Fuel cell energy management field

YUAN Pengbin, ZHAO Yunhua, ZHANG Weichong

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

Abstract: Revealing and researching the information phenomenon of discrete distribution is an important topic of the intelligence science. This study verify feasibility about applying Bradford law to patent through theory method, area method and image method based on Fuel cell energy management field, and conclude the characteristics of patent information distribution. Study finally concluded that the patent information as well as periodical information exists obvious phenomenon of the concentration and dispersion and even stronger. It is Feasible to use Bradford law to patent.

Keywords: Bradford law, patent, discrete distribution, area method, image method, Sri LankaMali Cove unified equation

布拉德福定律(简称“布氏定律”)按照位次/频率排序的方法, 揭示科技期刊中文献信息的离散分布^[1]。随后该定律多次运用到确定核心期刊、优化馆藏、制定采购策略、比较学科成熟度、评价检索系统等方面^[2]。

近几年, 有学者根据布氏定律的思路, 将其方法移植到专利信息的离散分布研究中, 用来确定核心专利分类号、核心专利权人等。例如, 张鹏等^[3]在2010年选取计算机领域中涉及“便携式终端上音像文件”领域的专利文献进行统计分

作者简介: 苑朋彬(1990—), 男, 中国科学技术信息研究所研究实习员, 研究方向: 技术竞争情报(通讯作者); 赵蕴华(1967—), 女, 中国科学技术信息研究所研究馆员, 研究方向: 重点科技领域研究; 张维冲(1991—), 男, 南开大学博士研究生, 研究方向: 文本挖掘。

基金项目: 国家科技支撑计划课题任务书“面向科技创新的专利信息加工与服务关键技术研究与应用示范”(2015XM56)。

收稿时间: 2017年9月15日。

析，初步探讨了布拉德福文献分散定律在专利应用中的可行性；吕义超等^[4]在2011年以混合动力汽车的技术主分类号进行统计分析，并根据维克瑞（B.C.Vickery）多区域划分结果进一步探讨了布拉德福专利应用的可行性；黄慕萱等^[5]对专利引文中的信息离散分布现象也进行了一定的探讨。在相关研究中同样存在一些不符合布拉德福划分的情况。胡晨希、邵蓉^[6]在2012年对中国药品专利进行了布氏定律划分，计算值与观察值出现了一定偏差。然而目前将布氏定律应用到分析专利文献的系统性研究较少。

本文在以上学者研究的基础上，选取了技术发展相对迅速，专利数量比较适中，生命周期过程较为完整的燃料电池汽车能量管理技术领域，分别从理论法、区域法和图像法3个方面进行可行性研究，并以布氏定律的数学公式计算值作为区域法和图像法的参考，考察了该技术领域专利信息与期刊信息离散分布现象的异同。

1 研究方法

布氏定律是基于期刊文献提出的，描述期刊论文离散分布状态。专利信息和期刊信息是两种不同的信息源，有其特有的著录特点，如一个技术分类号可以对应多项专利，而一篇论文仅可以对应一种期刊，专利的复杂性造成了数据间的冗余现象。尽管如此，专利信息与期刊信息一样存在明显的集中与分散现象，采用布氏定律思路分析专利信息离散分布现象的可行性主要体现在以下两个方面。

(1) 科学统一性决定专利文献具有多个技术类别。科学统一性揭示的是学科之间的交叉现象。技术的发展必定和其他技术的发展相关联^[7]，即科学技术中的每一个学科都在一定程度上与其他学科进行关联。从而造成一篇专利文献具有多个技术类别的情况。以专利号为CN105091016-A的中国专利为例，其国际专利分类（IPC）不仅可以分布在以B01D-053/62^[8]（碳氧化物）为代表的技术分支下，而且可以分布在以F01K-007/40（用两个或多个串联的给水加热器）为代

表的技术分支下^[9]。

(2) 专利文献普遍存在离散分布。文献信息离散分布揭示的是，大部分的相关信息集中在少量的期刊中，这是一种社会因素支配的情报学现象^[10]。而在专利技术中，由于技术发展水平和人为竞争意图等原因，IPC分类下的专利数量同样出现明显的集中、分散趋势，表现在处于萌芽期、成长期、成熟期和衰退期的专利技术IPC数量、种类的动态变化^[8]以及专利申请中技术主题的变化。

现采用布氏定律的区域法和图像法进行可行性验证。首先，选取了区域法中能够准确描述区域分布的Egghe系数算法以及图像法中适用性较好的斯马利科夫统一方程计算值作为参考标准，探讨专利信息离散分布的特点；其次，对燃料电池汽车能量管理技术领域的专利信息进行收集，并以“主IPC”作为划分角度，按对应专利数量的高低划分3个区域（第一区、第二区、第三区）；最后，对各区域的专利信息进行提取，与参考计算值进行对比分析。

2 数据来源及预处理

本实验数据选取DII数据库中收录的燃料电池汽车能量管理技术的相关专利。运用关键词加德温特手工代码进行检索，经查重后得到相关专利829条。采用“主IPC”模式分析时，由于一篇专利有多个IPC，而在特定国家申请时仅有一个主IPC，所以对德温特数据中的PD（Patent Detail）字段进行逐条处理，删除重复IPC，保留同族专利申请国的主IPC分类号。然后将处理好的数据按照位次/频率排序，形成布氏分布表（表1），其中 r 代表序列号， $R(r)$ 代表累计专利数。由于专利号与专利分类号是1:n的关系，因此 $R(r)$ 累计总数量大于实际总数量。

3 区域法分析

区域法的文字表述为：按照刊载某一学科领域相关文献的高低，将期刊划分为3个区域（核心区、相关区、外围区），可以发现基本等论文

量的3个区域内的期刊比值符合1:n:n²的比例关系。反映到专利文献中可以进一步表述为按照某一技术领域相关专利的高低排序将对应的“主IPC”分类号划分为3个区域，每个区域对应的“主IPC”分类号之间的比值也应该符合一定的比例关系。以Egghe博士布拉德福系数计算值作为参考^[9,11]，并对实际划分的“主IPC”数量比值进行计算，从而可以清楚地对比得出专利信息离散分布的特点，以及采用布氏定律思路运用到专利区域描述的可行性等问题。如表2是实际的“主IPC”分类号分布区域表；式(1)和式(2)是布拉德福系数计算公式；表3是具体的实际值与计算值之间的差异对比表。

将式(1)和式(2)应用到专利信息中，P代表分区数。本文仅讨论分区P=3的情况；K代表比例系数，Y_m为排名第一的IPC对应的专利数量；E为欧拉系数，e^E≈1.781。r₀为核心区包含的IPC分类号的数量，T代表专利分类号总数。

根据表1中的相关数据计算得出比例系数K和第一区的IPC数量r₀，最后根据r₀:r₁:r₂=1:n:n²的比例关系计算出各区的理论分布值，与表2相结合形成表3。

$$K = (e^E \times Y_m)^{\frac{1}{P}} \quad (1)$$

$$r_0 = T \times (k-1) / (k^3 - 1) \quad (2)$$

结合表1、表2、表3可知，燃料电池汽车能量管理技术领域主要包含392个“主IPC”分类号，经过区域划分以后实际得到的各区域的主分类号数量为8个、50个、334个；计算得到的各区域的主分类号数量分别为9个、54个、329个，实际值划分数量与计算值之间存在很小的差异，但可以明显地看出，第一区域和第二区域的“主IPC”分类号实际数量要小于计算值，第三区域的数量要大于计算值，这在一定程度上反映出两方面内容：第一专利信息的集中度要高于期刊文献，表现在第一区域和第二区域中，更少的“主IPC”分类号中包含了更多的专利文献；第二

表1 燃料电池汽车能量管理技术领域主IPC布氏排序表

r	专利量	主IPC	R(r)
1	128	B60L-011/18	128
2	72	H01M-008/04	200
3	58	H02P-027/06	258
4	45	B60L-003/00	303
...
392	1	H05K-007/02	1390

表2 燃料电池汽车能量管理技术领域主IPC布氏分布表

区域	分类号	主IPC量	专利量/件
第一区	B60L-011/18、H01M-008/04、H02P-027/06、B60L-003/00、H02J-007/00、B60L-015/20、H02P-021/00、H02P-006/08	8	475
第二区	H02M-007/48、H02M-003/155、H02P-006/00、B60L-001/00、H02P-027/04、H02P-029/00、B60L-015/00、B60W-020/00、H02P-023/00、H02P-027/08等	50	461
第三区	G06F-001/26、G06F-007/00、G06Q-010/00、H01M-010/42、H01M-010/46、H02H-007/08、H02J-003/38、H02M-001/08、H02P-003/18、H02P-006/06等	334	454

表3 燃料电池汽车能量管理技术专利信息离散实际值与计算值比较

区域	实际值		计算值	
	主IPC分类号	比例系数K	主IPC分类号	比例系数K
第一区域	8	—	9	—
第二区域	50	6.3	54	6.0
第三区域	334	6.7	329	6.1

区域专利信息的学科交叉型要大于期刊文献，表现在第三区域“主IPC”分类号数量和种类的扩张，这是专利信息离散分布相比期刊文献信息分布的一个特点。

比例系数K是反映信息离散分布的一个重要的参数，K值越大，核心效应越强，K值越小，核心效应越弱。理论计算值的比例系数均值为6.05，各区域间基本符合1:6:6²的布氏比例关系；而实际的比例系数均值为6.5，且各区域的关系也基本符合一定的比例关系，为1:6.3:6.7²，这进一步说明了专利信息存在着与期刊信息一致的集中分散趋势，将布氏定律区域法的研究思路运用到专利信息离散分布现象的研究在一定程度上是可行的。

4 图像法分析

图像法的文字表述为：以等级排列的期刊对应的序号Lg(r)为横坐标，以论文累计数量为纵坐标，可以将论文随期刊序号的变化情况在图表中表现出来，形成布拉德福分散曲线，曲线主要由3段组成，凹曲线、直线、凸曲线（不一定出现）；反映到专利信息中，可以进一步表述为以等级排列的“主IPC”对应的序号对数值lg(r)为横坐标，以专利累计数量为纵坐标，可将专利随期刊序号的变化情况在图表中表现出来。而图像法的证明中常采用一定的数学模型对其分散规律进行刻画，斯马里科夫统一方程是前苏联学者СМОЛИКОВ于1977年基于科技期刊论文提出的并被用来描述布拉德福分散曲线的数学方程之一^[12]，以此方程作为专利信息离散分布程度的参考，能够揭示专利信息的离散分布状况。同时，方程描述专利数据离散分布的好坏主要采用K-S检验的方法，K-S检验（Kolmogorov-Smirnov）^[13]又称D检验，是一种拟合优度检验的方法，常被用来检验一组样本数据的实际分布是否符合某一指定的理论分布。式（3）和式（4）分别描绘了斯马里科夫同一方程的基本转换公式；图1显示了实际值和理论计算值图像之间的差异；表4例举了K-S检验结果的相关数据。

$$R(n) = K \lg(r + pe^{-qr}) + b \quad (3)$$

$$R(x) = K \lg(e^x + pe^{-qe^x}) + b \quad (4)$$

其中，K、P、q、b为方程的参数，r为对应的专利序号，若令 $x=\ln(r)$ ，式（3）对应的方程则变成了关于x的方程，如式（4）。此方程可用一元非线性回归来拟合这四个参数，拟合效果如图1，拟合的基本方程式为：

$$R(x) = 233.673511 \lg(e^x + 0.70333e^{-0.01089e^x}) - 42.77234$$

如图1所示，线1代表实际分布曲线，线2代表理论拟合曲线，虚线代表直线部分的趋势线；观察实际分布曲线可知，专利信息的离散分布曲线主要包含两段曲线：初始缓慢增长的凹曲线部分和快速上升的直线部分，其直线和曲线部分的交点A值大约为2，其对应的核心区专利数量大约为8（计算方法为 e^A ），与上述区域法划分的第一区域的结果保持一致，这在一定程度上说明专利文献的图像分布同样符合布氏定律图像的描述。另外，观察线2的理论拟合值曲线可知，运用斯马里科夫统一方程描述专利信息离散分布的拟合效果较好，其差异主要体现在头部的拟合，考虑其主要原因在于斯马里科夫统一方程是基于期刊文献提出的，而专利文献的集中度较期刊文献的集中度要高，所以运用该方程描述专利信息会存在一定的差异；模型的拟合好坏主要通过K-S方法进行检验，通过实际值频数和理论值频数的对比，找出最大的差异点，然后参照抽样分布，确定此差异是否出于偶然，基本的检验结果如表4所示。

表4分别计算了实际累计频率和理论累计频率，然后根据 $D = |(Fn(x) - F0(x))|$ 找出最大临界值差值Max(D)。若 $\text{Max}(D) > D(n, \alpha)$ ，表示该专利信息离散分散图像分布不符合斯马里科夫统一方程的描述。反之则符合。其中， $D(n, \alpha)$ 是显著性水平为 α ，样本量为n的拒绝临界值，计算方法为 $1.36/\sqrt{n}$ 。该样本量（即为样本容量）是392。令 $\alpha=0.05$ 。表示模型在95%的置信区间内是可信的，因此计算得到 $D(392, 0.05)$ 为0.069，

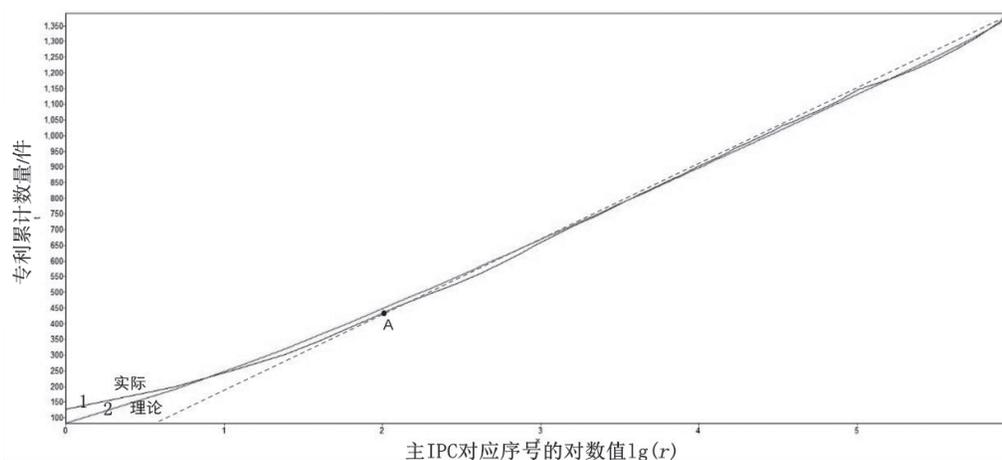


图1 燃料电池汽车能量管理技术领域的拟合效果图

表4 燃料电池汽车能量管理技术领域的K-S检验数据表

r	R(r)	实际累计频率Fn(x)	理论累计频率F0(x)	MAX(D)
1	128	0.0921	0.0595	0.0326
2	200	0.1439	0.1374	0.0065
3	258	0.1856	0.1904	0.0048
4	303	0.2180	0.2306	0.0126
5	347	0.2496	0.2631	0.0134
...
392	1390	1.0000	0.9934	0.0066

$\text{Max}(D) < D(392, 0.05)$ 。所以燃料电池汽车能量管理技术领域的专利信息离散分布现象符合斯马利科夫统一方程的描述。

图1以及方程拟合效果均有效地证明了专利信息存在着与期刊文献信息一样的集中与分散趋势,并且采用斯马利科夫统一方程能够较为精准地描述专利信息离散分布,这在一定程度上验证了将布氏定律的图像法研究思路应用到专利信息的离散分布中也是可行的。

5 结论

本研究选取燃料电池汽车能量管理技术专利信息数据,从理论可行性、区域可行性和图像可行性3个方面进行分析,结果显示燃料电池汽车能量管理技术专利信息的离散分布完全符合布氏定律。

(1) 燃料电池汽车能量管理技术专利信息存在着明显的集中与分散现象,其划分的各区域符

合一定的离散分布规律。

(2) 燃料电池汽车能量管理技术专利信息离散分布的特点是,第一区域和第二区域的集中度较强,第三区域的技术交叉扩散程度较广。

(3) 从燃料电池汽车能量管理技术专利信息的布氏区域划分可以看出,燃料电池汽车能量管理技术领域的核心专利技术分支是以H02P(电动机、发电机或机电变换器的控制或调节;控制变压器、电抗器或扼流圈)、B06L(电动车辆动力装置,车辆辅助装备的供电;电力制动系统、电气安全装置)、H01M(用于直接转变化学能为电能的方法或装置,例如电池组)、H02J(供电或配电的电路装置或系统;电能存储系统)等为代表的几大技术分支。

(4) 本文仅选取了生命周期发展过程较为完整的燃料电池汽车能量管理技术专利的数据对布氏定律进行了验证,其他各技术领域专利数据的离散分布是否符合布氏定律,有待后续研究,对

不同生命周期技术领域专利信息的离散分布作进一步分析。

参考文献

- [1] Vickery B C. Bradford law of scattering[J]. Journal of Documentation, 1948, 4(1): 199-203.
- [2] 张海燕. 关于布拉德福定律及其应用的若干思考[J]. 情报探索, 2013, 194(12): 19-21.
- [3] 张鹏, 刘平, 唐田田, 等. 布拉德福定律在专利分析系统中的应用[J]. 现代图书情报技术, 2010, 26(7): 84-87.
- [4] 吕义超, 刘红光, 王君. 布拉德福定律在专利文献中应用的可行性研究[J]. 情报分析, 2011, 4(2): 49-52.
- [5] Huang M, Huang W, Chang C, et al. The greater scattering phenomenon beyond Bradford's law in patent citation[J]. Journal of the Association for Information Science & Technology, 2015, 65(9):1917-1928.
- [6] 胡晨希, 邵蓉. 基于布拉德福定律的药品核心专利分

析[J]. 中国药事, 2012, 26(2): 134-136.

- [7] 邱均平. 信息计量学(四): 第四讲文献信息离散分布规律-布拉德福定律[J]. 情报理论与实践, 2000, 23(4): 315-320.
- [8] 苑朋彬, 赵蕴华, 周肖贝. 中美碳捕集与封存技术专利布局研究[J]. 全球科技经济瞭望, 2016, 31(2): 65-72.
- [9] 苑朋彬. 基于技术生命周期法进行技术预测研究: 以半导体光学光刻技术为例[D]. 北京: 中国科学技术信息研究所, 2016: 55-58.
- [10] 马费成, 陈锐. 科学信息离散分布的机理分析[J]. 中国图书馆学报, 2000, 26(5): 20-22.
- [11] 王崇德, 赵艳. 布拉德福分布解析式的择优评鉴[J]. 情报学报, 2010, 17(4): 2-8.
- [12] 邱均平. 信息计量学[M]. 武汉: 武汉大学出版社, 2007: 109-123.
- [13] 周爱民. 几种布拉德福分散曲线拟合模型的实证比较[J]. 情报杂志, 2013(1): 59-62.

(上接第 39 页)

提供企业之间的无缝衔接, 使得供应链的路径更短, 企业之间的衔接更加紧密。同时也通过企业之间的协作实现数据共享, 找到更加科学合理的生产运作方案。

(3) 构建协同创新平台在实现创新资源投入产出效益最大化的同时, 应当满足不同的创新主体及其应用目标, 因此, 对本研究提出的 3 种不同的实现路径和最优路径策略, 随着相关研究的不断深入, 需要结合当前已有的创新平台实践经验, 丰富平台实现路径, 提出更加实用、有效的平台路径策略。

参考文献

- [1] 张巍. 供应链企业间的协同创新模型研究[D]. 重庆: 重庆大学, 2009.
- [2] 王梓蓉. 知识链节点企业间的知识溢出过程与协同创新均衡研究[J]. 中国科技资源导刊, 2014, 46(1): 169-75. DOI: 10.3772/j.issn.1674-1544.2014.01.012.
- [3] 牛小格. 供应链协同创新环境下企业间的知识共享机制研究[D]. 天津: 天津大学, 2014.
- [4] 周华, 周水银. 基于顾客价值的供应链企业协同创新

研究[J]. 管理学报, 2016, 13(10): 1557-1562.

- [5] 陈大辉. 基于“互联网+”的协同创新平台构建[J]. 图书馆工作与研究, 2016(3): 16-19.
- [6] 王艳, 纪志成. 基于大数据的物联网产业协同创新平台研究[J]. 贵州社会科学, 2015(6): 139-143.
- [7] 王少华, 王克勤. 基于云计算的产学研协同创新平台研究[J]. 机械设计与制造工程, 2014(5): 1-4.
- [8] 黄南霞, 谢辉, 王学东. 大数据环境下的网络协同创新平台及其应用研究[J]. 现代情报, 2013(10): 75-79.
- [9] 高玲. 协同创新的平台与模式[J]. 科技与创新, 2016(14): 41.
- [10] 程强, 石琳娜. 基于自组织理论的产学研协同创新的协同演化机理研究[J]. 软科学, 2016, 30(4): 22-26.
- [11] 吕璞, 林莉. 基于开放式创新的供应链企业协同创新模型研究[J]. 科技管理研究, 2014(1): 197-200.
- [12] 潘郁, 路书星, 潘芳. 大数据环境下产学研协同创新网络生态系统架构[J]. 科技进步与对策, 2014(8): 1-4.
- [13] 赵龙文, 冯小宁. 基于 OGSA 的广东产业集群协同创新平台设计[J]. 科技进步与对策, 2010, 27(14): 32-35.
- [14] 严建援, 甄杰, 董坤祥, 等. 区域协同发展下创新资源集聚路径和模式研究: 以天津市为例[J]. 华东经济管理, 2016, 30(7): 1-7.
- [15] 王翔, 潘郁. 基于云计算的协同技术创新平台[J]. 计算机工程与应用, 2011, 47(15): 57-60.