

开放获取文献分布及其博弈分析

Open Access Documentation Distribution and its Game Analysis

崔景昌^{1,2} 刘德洪¹

(1. 中国科学院国家科学图书馆 北京 100080; 2. 中国科学院研究生院 北京 100049)

摘要 首先对开放获取的文献分布进行了统计分析,发现开放获取文献分布在不同学科领域不同地域范围存在着不均衡性。进而类比于经济上利益最大化原则,在设定科学研究人员的目的是最大化他们的名望和影响力的前提下,从博弈论的角度分析了不同学科开放获取文献分布不均衡的原因,并指出要促进开放获取必须加强从国家到机构等不同层面的引导,同时获得研究人员的合作。

关键词 开放获取 博弈分析 文献分布

近年来,科学交流模式和科学出版模式悄然地进行着演化,传统的出版机制正承受着越来越多的挑战。首先,数字化需求的增长促进了电子出版的发展;再者,进入 20 世纪 90 年代以来,期刊价格的上涨比例远超出同期图情机构预算经费的增长,使得图情机构不得不减少期刊的订购,而砍刊的结果就是造成期刊价格更高,形成恶性循环,因而更加剧了“期刊危机”(serials crisis);另外,由于商业出版商对学术出版的垄断,除了实行高价政策外,他们还对获取做出种种限制,只有在购买了许可后才可对文章进行利用,而这种做法又直接导致了“获取危机”(permissions crisis)^[1]。学术交流危机促使各国学术界、图书情报界以及出版界都开始积极探寻更有效的科学交流模式,开放获取(open access)正是在这样的背景下应运而生,同时信息技术及网络技术为开放获取提供了实现可能。

开放获取的宗旨在于利用网络条件实现科学研究成果的广泛共享,研究人员的非盈利的信息传播动机和利用需求是开放获取赖以存在的基本前提。从“布达佩斯倡议”到“柏林宣言”,从 arXiv.org 到 CiteBase、CiteSeer 等,开放获取已经广泛渗透到很多学科领域,越来越多的国家也都开始采取措施不同程度地促进开放获取。当前,实现开放获取主要有两种途径,一是“金色之路”(golden road)开放获取期刊,二是“绿色之路”(green road)机构知识库,前者是指作者直接将论文或研究成果投稿于开放获取期刊,后者是指作者在投稿于非开放获取期刊之外将研究成果提交至机构知识库^[2]。本文拟就开放获取现状,即文献分布等做一个统计分析,然后在此基础上从博弈论的角度分析造成这种分布的原因,并指出如何才能更有效地促进开放获取。

1 开放获取文献分布

开放获取经过十年来的发展,规模不断扩大,影响力也不断提高,为学术交流提供了新的途径,同时也极大地提高了学术交流的效率,加速了学术成果传播。其直接表现就是开放获取提高了论文的引用次数和即期影响因子,目前很多学者在不同学科领域的研究均证明了这一点。开放获取影响力的扩大,吸引了更多的机构和国家开始关注开放获取建设,那么我们就有必要研究开放获取文献的具体分布情况,从而更深层次地揭示当前开放获取的特征。

1.1 开放获取期刊与机构知识库 目前,据 DOAJ(Directory of Open Access Journals)^[3]的统计,它收录了 2700 多种同行评议的开放获取期刊,其中有 807 种期刊可以在文章层面进行检索,包括 134 264 篇文献,而 J-Gate^[4]收录的开放获取期刊达到了 3 896 种。全世界大约出版 2.5 万种科技与学术期刊,开放获取期刊在其中占的份额已经超过了 15%。另据 ROAR(Registry of Open Access Repositories)^[5]的统计,在该机构登记的机构知识库已接近 700 个,文章数接近 60 万篇。DOAR(Directory of Open Access Repositories)^[6]统计数字为 881 个。由此可见,对于可开放获取的文献来说,机构知识库要远大于开放获取期刊,即大部分的开放获取文章都存于机构知识库中,这也正是因为开放获取的绿色之路较金色之路实现起来要容易。

1.2 开放获取文献的学科分布 DOAJ 收录的开放获取期刊涵盖农业和食品科学、生物和生命科学、化学、历史和考古学、法律和政治学、语言和文学等 17 个学科和主题领域。其中医学类期刊 229 种,教育 185 种,计算机科学和数学均为

102 种,其它学科就相对较少。同样从机构知识库中文献的学科主题分布也能发现开放获取文献不同学科上差异是很大的, DOAR 统计的不同学科知识库占所有知识库的比例中,计算机和 IT、医药和卫生的比例达到了 8%,大部分学科的比例都在 1%~3% 之间。当然不同学科的规模与发展确实也客观地决定了比例,但开放获取在不同学科上的应用程度确实是不均衡的,有的学科几乎全部采用开放获取,而有的学科还几乎完全是传统的出版方式。

1.3 开放获取文献的地域分布 开放获取策源地以及主要的推动力量都在欧美,这两个地区的开放获取文献占全世界开放获取文献的 80% 左右,其它地区较前几年已经有了较为显著的增长,但开放获取地域分布上的不平衡还是相当的明显。据 DOAR 统计的各大洲开放获取机构知识库的分布数据显示,在所收录的 881 个知识库中欧美地区的达到了 718 个。按机构知识库数量排序,接下来的几个洲分别为大洋洲、亚洲、南美洲、非洲。按国别统计的分布情况为,美、德、英、澳、荷、法、瑞、加是机构知识库数量最多的 8 个国家,差不多占到了全部数量的 80% 左右。由此可见,尽管很多国家都开始重视开放获取,但其在不同国家的采用程度也是不均衡的。

2 开放获取文献分布的博弈分析

开放获取出版在不同学科中的实现程度是不一样的,以计算机、数学、物理和医药卫生等学科为主导,在这些学科里面形成了良好的开放获取氛围。以 arXiv 为例,它的主要用户群是物理学和数学的研究人员,这些领域的研究人员除按传统方式投稿到同行评议的期刊之前,往往都会自行把研究成果提交到 arXiv 服务器,近些年来这些领域很多重要的文章甚至都没有投稿到传统期刊中。相反,在很多其它学科领域里,尽管支持开放获取的呼声也很高,但却始终得不到广泛的利用,研究人员仍然投稿到传统的并不能免费获取的期刊。那么就产生了这样一个问题,既然都认为开放获取好,可是为什么开放获取只在少数几个领域应用的比较广泛呢?

本文拟从博弈论的角度进行一下分析,首先我们做一个假定,即科学研究人员的目标是最大化他们的名望,类似于经济学上把人都视为最大化自身利益的经济人。科学名望来源于两个途径,一个是该研究人员论文的被引用次数,另外一个则是他们所发表论文的期刊的影响力^[7]。已有的大量研究均表明了开放获取能够极大地提高论文的被引用次数。下文从两个研究人员在面临选择时将论文投诸于开放获取还是传统期刊的情境入手,展开两位研究人员的博弈。

2.1 初始化 首先,我们以标准化的格式定义两人的开放获取博弈为 Γ ,两位准备投稿的研究人员(研究人员 1 $\triangleq A$, 研究人员 2 $\triangleq B$),他们可以采取的策略为, $\Psi^A = \{s_1^A, s_2^A\}$, $\Psi^B = \{s_1^B, s_2^B\}$,两种策略分别是指作者选择开放获取(o)或选择传统期刊(ϕ),由此构造开放获取的博弈树如图 1 所示。

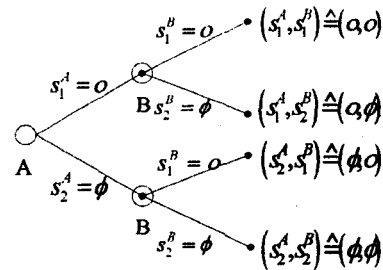


图 1 开放获取博弈树

整个博弈的策略组合是由两位准备投稿的研究人员所采取策略的笛卡尔积组成,即 $\Psi = \Psi^A \times \Psi^B = \{(o, o), (o, \phi), (\phi, o), (\phi, \phi)\}$

如前所述,假定研究人员的目的就是最大化他们的名望,为了更有效地进行分析,我们进一步设定这两个研究人员所处的科学团体还没有广泛地采用开放获取,并且传统的出版商会倾向于拒绝已经索引到开放获取服务器的文章。那么博弈的所有结果就如表 1 矩阵所示。

表 1 两人博弈所有可能的结果矩阵

| | | |
|-------|--------------------------------|-------------------------------|
| A \ B | o | phi |
| o | (gamma + delta, gamma + delta) | (gamma - alpha, gamma + beta) |
| phi | (gamma + beta, gamma - alpha) | (gamma, gamma) |

为了更容易比较,我们假定两位研究人员的初始名望是一样的,用 γ 来表示。如果两人都选择在传统期刊上发表论文,即采取(ϕ, ϕ)策略,他们的名望 γ 不发生变化。如果两人中有一人采取开放获取策略的话,即(o, ϕ)或(ϕ, o);变量 α 和 β 用以描述两人名望的减少量和增加量。通过这样建模,可以看出,当其中一个人采用开放获取,同时另一个不采用开放获取,前者的名望就会降低。这主要是因为非开放获取的科学团体内,研究人员的影响力和名望主要是由他们所发表文章的期刊的名望和影响力而决定的,如果投诸开放获取,研究人员就不能获得相关期刊的名望和影响力。通过设定 $\alpha = \beta$,名望就可看成是个相对量。变量 δ 表示两人都采取开放获取策略(o, o),每人的期望结果均为 $\gamma + \delta$,在这种情况下,所有研究人员均采取开放获取策略,出版商就不得不接收已经可开放获取的论文,研究人员不但获得了期刊带给他们的名望,而且通过开放获取获得了更多的引文数。最后一种情况当然是最理想的状态了。那么在具体的情况下,结果到底如何下面将描述三种具体变量设定下的开放获取博弈。

2.2 零和博弈(zero sum game) 将博弈结果集中的自由变量设定为 $\gamma = 0, \delta = 0, \alpha = \beta = 1$,即是最简单的一种博弈情况。两人的初始名望以及采取开放获取所带来的名望均设定为 0, $\alpha = \beta$ 。这种设定将名望设定为相对量,一人名望的增加必然导致另一人名望的等量减少,因此在这种情形下, δ 必然为 0,这样才不会导致整个系统名望量的增加,具体的结果矩阵见表 2。

表 2 将名望设定为相对量情况下的博弈结果矩阵

| A \ B | o | ϕ |
|--------|--------|--------|
| o | (0,0) | (-1,1) |
| ϕ | (1,-1) | (0,0) |

在这个博弈中,每个人的占优策略都是(ϕ),纳什均衡是(ϕ, ϕ),因而在这种情况下,不会有任何参与者会选择开放获取。

2.3 囚徒困境(prisoners' dilemma) 这个博弈类似于经典的囚徒困境。名望在这里不是完全的相对量,如果两位研究人员均采用开放获取策略,名望将增加 δ ($\delta > 0$)。在这种情况下,变量设定如下: $\gamma = 3, \delta = 1, \alpha = \beta = 2$ 。表 3 为两人的结果矩阵。

表 3 囚徒困境博弈下的结果矩阵

| A \ B | o | ϕ |
|--------|-------|--------|
| o | (4,4) | (1,5) |
| ϕ | (5,1) | (3,3) |

尽管两位研究人员如果都选择开放获取的话,他们的结果会更好,但在他们不明确对方选择的情况下,每个人最佳的选择只能是(ϕ),他们还是会采取传统的策略(ϕ, ϕ)。这种情况刚好描述了很多学科领域内存在的矛盾情形,研究人员一方面意识到如果采取开放获取的话他们本身将受益,但另一方面却没有人倾向于改变现状。

2.4 猎鹿博弈(stag hunt game) 猎鹿博弈描述的是两个猎人狩猎的情形,他们有两个选择,一个是猎鹿,一个是猎兔子,如果猎鹿的话必须两个人同时去,猎鹿会为两个人带来最大的利益,猎兔子不需要合作,每个人单独都可完成。在这里,开放获取类似于猎鹿,传统的出版方式类似于捕兔子。我们设定相关变量如下: $\gamma = 3, \delta = 1, \alpha = 2, \beta = 0$,结果矩阵如表 4 所示。

表 4 猎鹿博弈下的结果矩阵

| A \ B | o | ϕ |
|--------|-------|--------|
| o | (4,4) | (1,3) |
| ϕ | (3,1) | (3,3) |

与前两种设定不同的是,该设定有两个纯粹策略纳什均衡(pure Nash equilibrium), (o, o) 和 (ϕ, ϕ), 以及一个混合策略纳什均衡(mixed Nash equilibrium), $\frac{2}{3}(o, o)$ 。当(ϕ, ϕ)是风险占优的纯粹纳什均衡时, (o, o)是占优的结果。 $\frac{2}{3}(o, o)$ 表示当其中一个参与者认为另一方选择传统方式的概率大于 $\frac{2}{3}$ 时,他也会倾向于选择传统方式而非开放获取。

2.5 博弈分析及应对策略 零和博弈,囚徒困境是竞争性博弈,猎鹿博弈属合作性博弈。在本文开放获取的零和博弈中,两位研究人员由于面临此消彼长的选择,在不明确对方选择前提下,必然选择保守的行动,使得开放获取没有成长空

间。造成这样结果的原因是相关国家、学科领域、出版商等排斥开放获取,单一的作者如果选择开放获取非但没有好处,而且会使个人利益受到损害,那么就处于弱势的作者来说,他必然选择传统的出版方式,以保证自己的名望和影响力。

囚徒困境中,两位研究人员已经能够意识到开放获取会给他们带来好处,但由于在两者做出的选择不一致时,选择开放获取的一方会有一定的损失,在相互竞争中处于劣势,因而最终纳什均衡还是会停留在传统方式上。这种情形与零和博弈比较起来,已经有了一定突破,两位研究人员不完全是竞争关系,并且大环境也让他们知道了开放获取确有好处,并且是最佳结果,但由于限制因素的存在,使得他们做出保守选择。个人的理性导致了集体的非理性。猎鹿博弈引入了合作机制,出现了三个纳什均衡点,与前两者比较起来更为优化,充分意识到开放获取的重要性,加强前期的引导与合作,开放获取就会变得容易起来。但如果少了前期的引导与参与者之间的合作,人们选择开放获取的概率就会变小很多。

3 小结

开放获取为解决“学术交流危机”而生,如今已发展成一种新的出版方式和科学交流方式,已扩展到很多学科领域,很多国家也纷纷签署开放获取方面的协议以融合到开放获取的潮流中。但统计表明,开放获取依然还存在着学科分布、地域分布的不平衡性,当然这种不平衡和学科发展以及地区科技发展有直接关系,但我们不能忽视非此类因素的存在。三种博弈分别不同程度地模拟了开放获取在实行中受限制的情形,以及研究人员在相应大环境下所做出的选择。从中我们可以认识到,要让开放获取充分地发展,就必须改变游戏规则,要从不同层面上放宽对开放获取的限制,构建适合开放获取的大环境,加强引导,加强研究人员在意识上的合作,必将能使开放获取渗透到各个学科,以实现开放获取出版分布的相对平衡。

参考文献

- 1 初景利. 开放获取的发展与推动因素[J]. 图书馆论坛, 2006(6): 238-242
- 2 Tim Brody, Stevan Harnad, et al. The Access/Impact Problem and the Green and Gold Roads to Open Access[EB]. <http://users.ecs.soton.ac.uk/~harnad/Temp/impact.html>. 2007-5-20
- 3 The Directory of Open Access & Hybrid Journals[EB]. <http://www.doaj.org>. 2007-5-28
- 4 J-Gate[EB]. <http://www.j-gate.informindia.co.in/>. 2007-5-28
- 5 ROAR(Registry of Open Access Repositories)[EB]. <http://roar.eprints.org/index.php>
- 6 OpenDOAR[EB]. <http://www.opendoar.org/>. 2007-5-28
- 7 M Hanauske, et al. Quantum Game Theory and Open Access Publishing, Physica A (2007)[EB]. doi:10.1016/j.physa.2007.04.12

(责编:愚阳)