

基于差异性视角的期刊区分性测度及分析研究

张宝隆^{1,2}, 王 昊^{1,2}, 邓三鸿^{1,2}, 苏新宁^{1,2}

(1. 南京大学信息管理学院, 南京 210023; 2. 江苏省数据工程与知识服务重点实验室, 南京 210023)

摘 要 当前的期刊评价研究主要集中在影响力、声誉、质量等方面, 针对期刊内容差异性的定量研究相对较少。本文从差异性视角来对期刊进行评价分析, 提出期刊区分度指标来对期刊内容的差异程度进行度量。以 LIS (library and information science) 等 5 个学科的各 20 种核心期刊为研究对象, 首先对 LIS 期刊的内容差异性进行定量分析与评价, 然后从时间维度对 LIS 学科期刊区分度的变化规律进行了探测, 最后从学科维度对不同学科期刊个体和总体区分度的特征进行了分析和探讨。实验结果表明, 利用该指标能够很好地度量期刊研究内容差异性, 期刊区分度在时间维度上呈现出明显的变化规律, 不同学科期刊个体和总体的区分度均具有显著的学科特征。

关键词 期刊区分度; 差异性; 期刊评价; 层次聚类分析; PCA 降维

Measurement and Analysis of Journal Discriminative Capacity Based on Difference

Zhang Baolong^{1,2}, Wang Hao^{1,2}, Deng Sanhong^{1,2} and Su Xinning^{1,2}

(1. School of Information Management, Nanjing University, Nanjing 210023;

2. Jiangsu Key Laboratory of Data Engineering & Knowledge Service, Nanjing 210023)

Abstract: Current research on journal evaluation focuses primarily on measurement of influence, reputation, quality, and similar concepts. This paper evaluates journals from a new perspective based on content difference, proposing journal discriminative capacity as an index to measure differences in academic journals' content. Twenty core journals for each of the five disciplines, which are library and information science (LIS), aerospace, biology, art and law, were selected as research objects. First, content differences between LIS journals were quantitatively analyzed and evaluated. Trends in LIS journals' discriminative capacity over time were then explored. Finally, the characteristics of individual and overall discriminative capacity of journals from different disciplines were analyzed and discussed. The results show that the index is highly effective in measuring differences in journal research content, that journals' discriminative capacity shows obvious trends over time, and that individual and overall journal discriminative capacity has significant disciplinary characteristics.

Key words: journal discriminative capacity; journal difference; journal assessment; hierarchical clustering analysis; PCA dimension reduction

1 引 言

学术期刊评价长期以来都是图书情报学界研究

的重点问题之一, 对于学术交流和科研人员、研究机构的评估等具有重要的影响^[1]。目前, 学术界针对期刊评价的研究主要集中在学术期刊的影响力、

收稿日期: 2019-03-19; 修回日期: 2019-12-10

基金项目: 国家社会科学基金重大招标项目“情报学学科建设与情报工作未来发展路径研究”(17ZDA291); 国家自然科学基金青年科学基金项目“面向学术资源的 TSD 与 TDC 测度及分析研究”(71503121)。

作者简介: 张宝隆, 男, 1991 年生, 博士研究生, 研究方向为信息分析与科学计量; 王昊, 男, 1981 年生, 教授, 博士生导师, 研究方向为信息分析与科学计量, E-mail: ywhaowang@nju.edu.cn; 邓三鸿, 男, 1974 年生, 教授, 博士生导师, 研究方向为科学计量、学术评价; 苏新宁, 男, 1955 年生, 教授, 博士生导师, 研究方向为信息分析与科学评价等。

声誉、质量等方面，所采用的实验数据多为统计数据、引文数据等外部数据，没有充分考虑到期刊自身研究内容的定量分析。同时，当前一些学术期刊的研究内容“同质化”现象趋向明显，这在一定程度上也阻碍了学术研究的发展和创新。而已有的期刊评价研究鲜有涉及这些问题，因此，从内容差异性角度来评价期刊是非常必要且有意义的，并且如何对差异性进行度量也是一个非常有价值的研究问题。

在信息检索领域，有学者提出术语区分模型^[2-3]来对索引术语的质量进行定量评价，其核心是通过对文档内容的度量来分析术语在索引方面的差异性^[4]。虽然该模型计算复杂度较高，不适用大规模数据和更大粒度的学术对象，并且对术语差异性测度效果不太理想，但其研究思想则非常适用于内容差异性探究，这为期刊研究内容的定量评价提供了理论基础。因此，笔者将术语区分模型的研究方法扩展至学术期刊差异性评价，通过对研究内容的度量来揭示期刊的差异性特征，挖掘独具特色的期刊个体，从而为指导学术期刊的多向性发展、改变期刊学术研究“同质化”趋向提供参考依据。

本文在术语区分模型的基础上，提出“期刊区分度（journal discriminative capacity, JDC）”，其被定义为：某种期刊与指定范围内其他所有期刊之间的总体差异性及其程度；并给出测算方法来度量学术期刊研究内容之间的差异程度，JDC越大，说明该期刊较其他期刊的差异越大，研究内容更独特、更具个性，反之则说明该期刊可能具有“同质化”趋向。因此，利用期刊区分度可以测算和分析期刊群体中不同个体之间的差异程度，挖掘独具特色或具有“同质化”趋向的期刊个体，从而实现从内容差异性这一新的视角来对学术期刊进行评价。为此，本文以期刊内容中的题名、关键词、摘要等作为实验语料，通过测算 LIS（library and information science）期刊的 JDC 来对期刊个体的差异程度进行分析和探讨，并在此基础上分别从时间维度和学科维度来探索 JDC 的时间特征和学科特征，以揭示不同时期学术期刊差异性的演化规律和不同学科期刊的差异化特征，为期刊的差异性评价和个性化发展提供理论指导。

2 近期相关研究

目前，针对学术期刊的评价形式多样，评价的

侧重点主要集中在期刊的质量、声誉、影响力等方面，所采用的方法主要包括同行评议法、科学计量法等^[5]。其中，同行评议法是早期的期刊评价方式^[6]，在学术评价中应用非常广泛，但由于同行评议可能存在个人主观性强、学术道德不规范等问题^[7]，该方法在期刊评价中逐渐被计量方法所取代。科学计量法是期刊评价方法中应用最为广泛的方法之一，主要采用单一指标或复合指标来对期刊进行定量评价，常见的指标有影响因子及衍生指标^[8-11]、H 指数^[12]及相关指标^[13-16]、类 PageRank 指标^[17-19]、综合评价指标^[20-22]等。当然，也有学者从其他视角对期刊进行评价，例如，基于知识积累角度的期刊质量评价^[23]、期刊竞争压力评价^[24]，等等。

针对差异性相关的研究主要侧重于学术研究对象差异性评价，例如，论文影响力评价指标的差异性评价^[25-26]，不同学科期刊评价指标的差异性分析^[27]，论文和综述中参考文献量的差异性评价^[28]，会议论文的领域与载体差异性^[29]，学生、教师、研究人员所发表的文章在期刊和年份中的差异性^[30]，国内外期刊在研究主题等方面的差异性^[31]，等等。这类研究主要是采用统计数据对不同学术对象之间的差异性进行对比分析，其中针对评价指标的差异性研究偏多。此外，也有一些学者从检索术语的差异性、研究主题的差异性^[32]、期刊选题的差异性^[33]、期刊发文的差异性^[34]等方面进行评价研究，虽然这些研究都是从内容差异性进行研究，但它们主要是针对较小粒度的术语、主题、论文等研究对象，所采用的方法难以适用于更大粒度的学术期刊对象。

通过对相关研究成果的梳理发现，针对期刊评价的研究多是利用各类指标来对期刊的质量、影响力等进行评价分析，很少有研究涉及期刊内容的差异性评价。在差异性研究方面，大部分研究集中在评价指标等研究对象的差异性对比分析，所采用的计量数据也多为引用量、发表量等^[35]外部统计数据；而针对术语、主题等小粒度学术对象的差异性研究方法也无法直接应用于期刊对象。所以，从整体上看，当前利用自身内容来探究期刊差异性的研究成果还不够丰富，已有的方法或指标难以对期刊的差异性进行度量。因此，本文创新性地提出 JDC 指标来对期刊内容差异性进行分析评价，并给出相应的测算方法使其更具操作性和实践意义，为学术期刊评价方法体系的发展和完善提供了新的思路。

3 数据和方法

3.1 数据来源和预处理

“中文社会科学引文索引”(CSSCI)和《中文核心期刊要目总览(2014年版)》(简称《总览》)均收录大量的精品期刊且进行了学科分类,所以,笔者按照期刊数量相当、被收录状况稳定的原则,分别从CSSCI和《总览》获取图书情报与档案管理、艺术学、法学、航空航天、生物科学这5个学科的期刊作为研究对象。然后根据CNKI数据库中

期刊的综合影响因子获取排名前20位的期刊作为实验对象,但有个别期刊部分年份的载文量过少,不足50篇,因此按照排名向下顺延的原则进行替换,最终所选期刊如表1所示。

为保证数据来源的一致性和数据质量,从CNKI数据库中获取实验数据。通常认为,标题、摘要、关键词能够很好地揭示论文研究内容,与全文相比,信息密度大、数据体量小。因此,笔者获取了2013—2017年5个学科共100种期刊的题录信息,将标题、摘要、关键词这3个字段作为本文的数据源,并对数据进行预处理:①题录数据的清洗

表1 5个学科期刊列表

图书情报与档案管理		航空航天		生物科学		艺术学		法学	
编号	期刊名称	编号	期刊名称	编号	期刊名称	编号	期刊名称	编号	期刊名称
1	档案学通讯	21	北京航空航天大学学报	41	古生物学报	61	北京电影学院学报	81	比较法研究
2	档案学研究	22	导弹与航天运载技术	42	基因组学与应用生物学	62	北京舞蹈学院学报	82	当代法学
3	大学图书馆学报	23	飞行力学	43	生命科学	63	当代电影	83	法律科学
4	国家图书馆学刊	24	固体火箭技术	44	水生生物学报	64	电影艺术	84	法商研究
5	情报科学	25	航空动力学报	45	生态学报	65	建筑学报	85	法学
6	情报理论与实践	26	航空学报	46	生态学杂志	66	美术	86	法学家
7	情报学报	27	航天返回与遥感	47	生物多样性	67	美术研究	87	法学论坛
8	情报资料工作	28	航天控制	48	生物工程学报	68	民族艺术	88	法学评论
9	情报杂志	29	航天器工程	49	生物化学与生物物理进展	69	南京艺术学院学报(美术与设计版)	89	法学研究
10	数据分析与知识发现	30	空间科学学报	50	生物技术	70	文艺研究	90	法制与社会发展
11	图书馆	31	空间控制技术与应用	51	生物技术通报	71	戏剧艺术	91	华东政法大学学报
12	图书馆建设	32	空气动力学学报	52	微生物学报	72	新美术	92	环球法律评论
13	图书馆论坛	33	南京航空航天大学学报	53	微生物学通报	73	戏曲艺术	93	清华法学
14	图书馆学研究	34	燃气涡轮试验与研究	54	微生物学杂志	74	艺术百家	94	现代法学
15	图书馆杂志	35	实验流体力学	55	遗传	75	艺术设计研究	95	行政法学研究
16	图书情报工作	36	推进技术	56	应用生态学报	76	音乐研究	96	政法论丛
17	图书情报知识	37	宇航材料工艺	57	应用与环境生物学报	77	音乐艺术(上海音乐学院学报)	97	政法论坛
18	图书与情报	38	宇航学报	58	中国科学:生命科学	78	中国电视	98	中国法学
19	现代情报	39	中国惯性技术学报	59	中国生物工程杂志	79	中国音乐学	99	中外法学
20	中国图书馆学报	40	中国空间科学技术	60	中国生物化学与分子生物学报	80	中央音乐学院学报	100	政治与法律

注:按照《总览》的学科分类,生物科学选取除植物学、动物学/人类学以外的期刊;CSSCI采用2017—2018版。

及抽样。原始题录中存在摘要、关键词等字段空缺的情况，这些题录可以视为无效数据进行剔除。由于期刊的出版周期、每期页数等不同，导致不同期刊的年载文量差距较大，最少的不足50篇，而最多的高达800余篇。因此，笔者采用分层随机抽样的方式从每个期刊的有效数据中每年随机抽取40条作为实验样本，以消除这些出版属性对实验结果的潜在影响。②文本分词及剔除停用词。首先采用中国科学院的NLPIR汉语分词系统对题录中的标题、关键词和摘要进行分词，然后结合哈工大停用词表、四川大学机器智能实验室停用词库、百度停用词表等对分词结果进行停用词过滤，将过滤后的字或词视为术语。值得注意的是，笔者将分词结果中的外文字符串如“iSchool”、“citespace”等视为术语，予以保留。

3.2 JDC计算方法

JDC主要用于描述不同期刊在研究内容方面的差异性及其程度，其核心思想是通过计算期刊空间(journal space, JS)中某一期刊被剔除之后期刊空间密度(journal space density, JSD)的变化来衡量该期刊的区分度。所以为了度量这一变化，需要首先构建JS，然后通过JSD的变化来计算JDC。具体计算方法如下。

1) JS构建

本文将由期刊-期刊矩阵(journal-journal matrix, JJM)构成的高维空间定义为JS，形如式(1)：

$$JJM = \begin{pmatrix} S_1 \\ S_i \\ \vdots \\ S_m \end{pmatrix} = \begin{pmatrix} s_{11} & \cdots & s_{1j} & \cdots & s_{1m} \\ \cdots & \cdots & s_{ij} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ s_{m1} & \cdots & \cdots & \cdots & s_{mm} \end{pmatrix} \quad (1)$$

其中，JJM是 $m \times m$ 维的对称矩阵； m 表示期刊空间中的向量个数，也即为期刊数量； $S_i = \{s_{ij} | i, j \in m\}$ 表示期刊向量； s_{ij} 表示期刊 i 与期刊 j 的余弦相似度，计算方法为

$$Similarity(C_i, C_j) = s_{ij} = \frac{\sum_{t=1}^n c_{it} \times c_{jt}}{\left(\sum_{t=1}^n c_{it}^2 \times \sum_{t=1}^n c_{jt}^2 \right)^{1/2}} \quad (2)$$

其中， C_i 和 C_j 表示术语向量，所有术语向量构成期刊-术语矩阵(journal-term matrix, JTM)，形如式(3)：

$$JTM = \begin{pmatrix} C_1 \\ C_i \\ \vdots \\ C_m \end{pmatrix} = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \cdots & \cdots & c_{in} \\ \cdots & \cdots & \cdots \\ c_{m1} & \cdots & c_{mn} \end{pmatrix} \quad (3)$$

其中，JTM为一个 $m \times n$ 维的矩阵； c_{it} 和 c_{jt} 为术语权重值； m 为术语向量的总个数（一个术语向量表示一个期刊）； n 为 m 个期刊所包含的术语总量。该矩阵是由<期刊编号，术语，术语权重>三元组转化而来，其中，术语权重表示方法有Presence（即出现为1，不出现则为0）、TF、TF-IDF，如公式(4)所示，

$$TF-IDF = TF \times IDF = tf_i \times \lg(m/m_i + \varepsilon) \quad (4)$$

其中， tf_i 为术语 t 在期刊 i 中出现的频率； m_i ($m_i \neq 0$)为术语 t 在术语向量中出现的频次； ε 为调节因子。本文中令 $\varepsilon=1/m_i$ ，以确保当 $m=m_i$ 时IDF不为零。

考虑到不同字段的术语对研究内容的揭示程度不同，需要对术语在不同字段中出现的频次进行加权，加权后的TF和TF-IDF分别记为TFw和TFw-IDF，其计算公式为

$$TFw = \frac{(\alpha n_{th} + \beta n_{tk} + \theta n_{ta})}{\sum_{i=1}^m (\alpha n_{th} + \beta n_{tk} + \theta n_{ta})} \quad (5)$$

其中， n_{th} 、 n_{tk} 、 n_{ta} 分别为术语 t 在期刊 i 的标题、关键词、摘要中出现的频次； α 、 β 、 θ 分别为术语在不同字段中的权重，本文中按照这3个字段术语的重要程度令 $\alpha : \beta : \theta = 4 : 2 : 1$ 。

2) JSD计算

本文将所有期刊向量与JS中心的平均相似度定义为JSD，其反映了期刊集合在空间中的聚集程度。这里的相似度采用基于距离的相似度进行测算，而这个距离则通过期刊向量 S_i 到JS中心的cosine距离来衡量，计算方法为

$$Dist(S_i, Centroid) = \left(\sum_{j=1}^m \left(s_{ij} - \frac{1}{m} \sum_{j=1}^m s_{ij} \right)^2 \right)^{1/2} \quad (6)$$

其中，Centroid表示JS中心，计算方法为

$$Centroid = \left(\frac{1}{m} \sum_{j=1}^m s_{1j}, \frac{1}{m} \sum_{j=1}^m s_{2j}, \cdots, \frac{1}{m} \sum_{j=1}^m s_{mj} \right) \quad (7)$$

因此，期刊向量 S_i 与JS中心的相似度就可以采用基于距离的相似度进行计算^[36]，

$$DS(S_i) = \frac{1}{c^{Dist(S_i, Centroid)}}, c = 1.3 \quad (8)$$

其中， c 是一个常量，它影响着距离对相似度的作用程度，当 $c=1.3$ 时，相似度结果最为理想。

进而，利用平均相似度来计算JSD，

$$JSD = \left(\frac{1}{m} \sum_{i=1}^m DS(S_i) \right) \quad (9)$$

3) JDC度量

JS中某一期刊被剔除后会形成新的JS，而其

JSD就会发生变化。因此,通过测算这种密度变化可以反映出该期刊与其他期刊的差异程度,计算方法为

$$JDC_{S_k} = \frac{JSD_{S_k} - JSD}{AVG_JSD} \quad (10)$$

其中, JSD_{S_k} 表示期刊 S_k 被剔除后的计算所得的 JSD, 计算方法同式(9); AVG_JSD 表示该 JS 的平均密度差, 主要用于密度差的标准化, 其中通过对密度差取绝对值来避免正负差值相互抵消, 计算方法为

$$AVG_JSD = \frac{1}{m} \sum_{k=1}^m |JSD_{S_k} - JSD| \quad (11)$$

4 实验结果及分析

4.1 LIS 学科期刊的 JDC 计算与分析

通过对 LIS 期刊 2017 年、2015—2017 年、2013—2017 年题录数据的预处理, 分别获得 5525、9203、11557 个术语, 以此为基础数据分别进行 JDC 测算。首先通过不同术语权重的合理性分析来设置术语权重, 然后对 LIS 学科期刊的内容差异性进行度量和评价。

4.1.1 术语权重的设置

为了确定最为合理的期刊的描述因子——术语权重, 笔者在计算术语权重的基础上将三元组转化为对应的 JTM (20×5525 维), 然后采用层次聚类法对 5 种 JTM 进行聚类, 结果如图 1 所示。

从 5 个聚类结果来看, 图 1a 中《图书馆杂志》(15) 被单独划分为一类, 其离散程度超过了档案学期刊, 这不符合常理, 表明 Presence 难以准确描述期刊内容; 在图 1b、图 1c 中《中国图书馆学报》(20) 与情报学期刊聚为一类, 而在图 1d、图 1e 中则刚好相反, 但其作为图书馆学领域的权威期刊, 与图书馆学期刊聚为一类则更为合理; 并且图 1b、图 1c 中各个簇的组间距离相对较小, 说明期刊之间的差异性没有被充分表现出来。所以, 图 1d 和图 1e 的聚类结果相较于其他更为合理, 且在图 1d、图 1e 中《图书馆学研究》(14) 与《图书情报工作》(16)、《图书情报知识》(17) 这种典型图书情报综合类期刊也聚为一类, 这也印证了这两种结果的合理性。

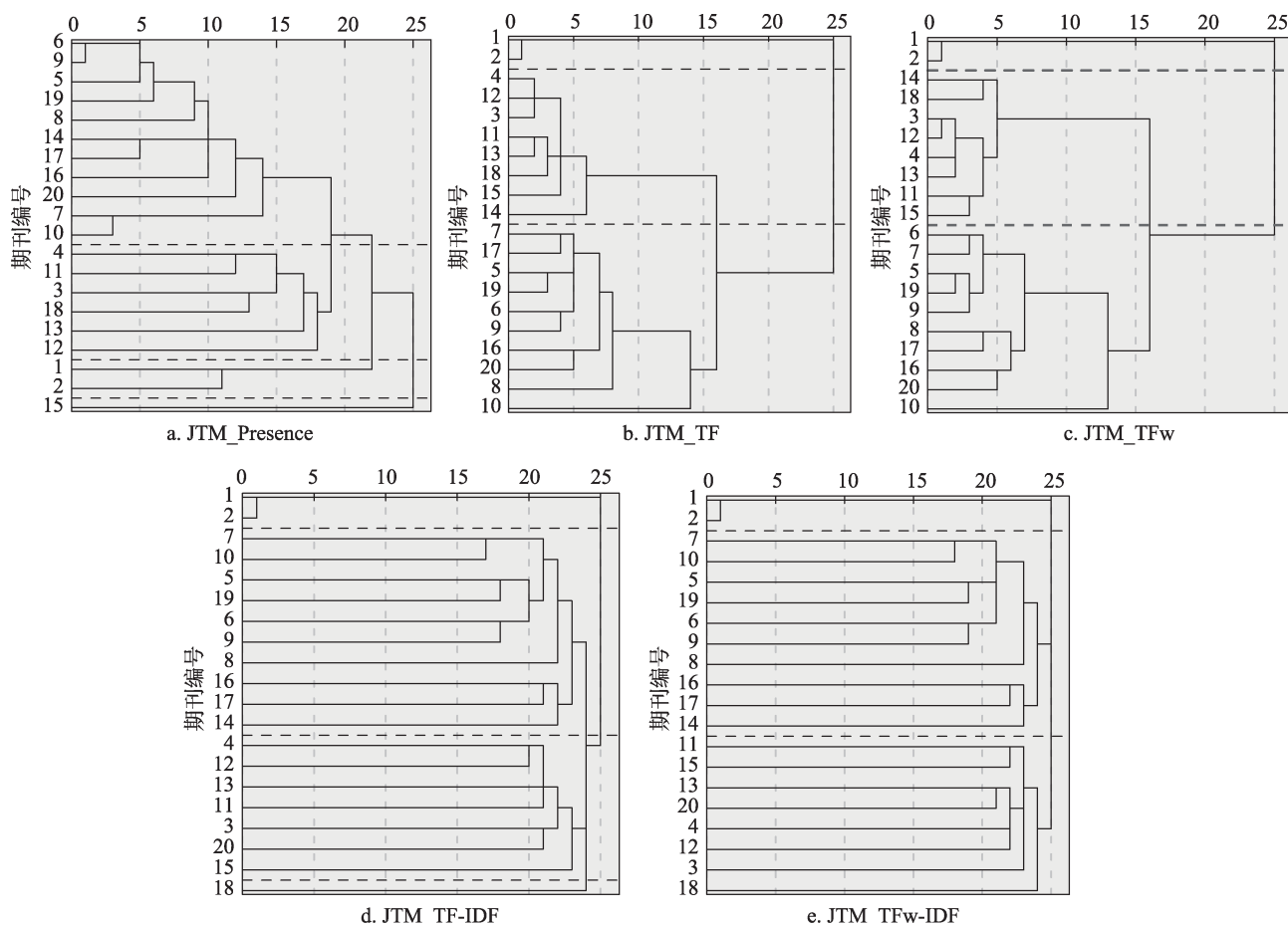


图1 基于5种术语权重的期刊聚类结果

因此，可以认为 TF-IDF 和 TFW-IDF 可能更加适用于 JDC 的测算。当然，这仅是基于聚类结果的初步判断，还需要进一步验证。笔者对 2017 年 LIS 期刊 JDC 进行了计算，结果如表 2 所示，5 种 JDC 结果在数值分布上存在明显差异。为了更加直观地观察这种差异，将表 2 中的数据绘制成箱线图，如图 2 所示。

表 2 基于 5 种术语权重的 JDC 结果

编号	Presence	TF	TFw	TF-IDF	TFw-IDF
1	2.79	3.13	3.01	6.99	7.30
2	1.01	2.84	2.85	6.94	7.29
10	4.12	1.86	1.68	0.94	0.82
7	1.01	0.84	0.89	0.50	0.44
15	3.00	0.21	0.20	0.50	0.49
6	0.33	0.87	0.95	0.45	0.42
18	0.26	0.43	0.29	0.44	0.40
4	0.30	1.78	1.75	0.42	0.46
5	-0.23	0.53	0.55	0.35	0.31
13	1.49	1.18	1.30	0.34	0.33
9	-0.23	0.77	0.79	0.31	0.23
12	1.56	1.66	1.56	0.27	0.16
3	0.72	1.50	1.66	0.15	0.26
11	0.91	0.36	0.32	0.13	0.28
8	-0.19	-0.16	-0.20	0.04	0.12
19	-0.33	-0.29	-0.28	-0.12	-0.13
16	-0.33	-0.54	-0.61	-0.17	-0.12
14	-0.77	-0.36	-0.38	-0.25	-0.04
17	-0.06	0.06	0.07	-0.27	-0.10
20	-0.36	-0.65	-0.66	-0.41	-0.28

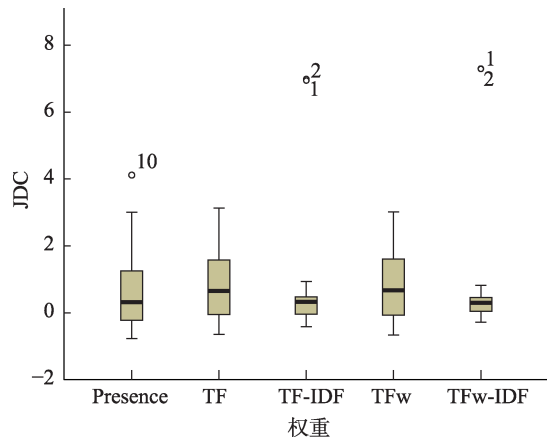


图 2 基于 5 种术语权重的 JDC 数值分布

观察图 2 可以发现，5 种 JDC 的数值分布具有明显差异。①基于 TF-IDF 和 TFW-IDF 的箱线图中出现了《档案学通讯》(1) 和《档案学研究》(2) 两

个孤立点，而其他期刊的 JDC 值分布比较集中，表明档案学期刊的研究内容与其他期刊差别很大，这比较符合当前图情档学科中档案学研究相对独立的现状。②基于 Presence 的箱线图中出现了《数据分析与知识发现》(10) 一个孤立点，这与当前 LIS 学科发展现状不符，也与其对应的聚类结果不一致，这也印证了 Presence 不适用于 JDC 计算。③基于 TF 和 TFW 的 JDC 均没有出现异常点，说明这两种术语权重没有充分体现期刊研究内容的差异性。所以，综合看来，基于 TF-IDF 和 TFW-IDF 的 JDC 结果最为合理，它们均可以作为术语权重应用于 JDC 算法，但考虑到算法的简便性，笔者在后续研究中将采用基于 TF-IDF 的 JDC 结果对期刊进行评价分析。

4.1.2 基于 JDC 的期刊差异性分析

将表 2 中基于 TF-IDF 的 JDC 结果绘制成散点图 (图 3)，同时，采用 PCA (principal component analysis) 法将 JS 映射到三维空间中 (如图 4 所示，图中黑点为期刊空间中心)，从而能够通过观测期刊的空间聚集状况来进行辅助分析。

观察图 3 可以发现，大部分的 JDC 都为正值，说明 LIS 期刊大都有一定的研究特色，在群体中具有较好的区分性。

(1) LIS 期刊的 JDC 出现了明显的分层。档案学期刊 JDC 最大值处于第一层，其余 JDC 为正、为负的期刊分别处于第二、三层。这与图 4 的空间分布一一对应，分别分布在空间的边缘 (区域①)、外围 (区域②)、中心区域 (区域③)。由此可以看出，档案学期刊的区分能力最强，与其他期刊差异最大，这也反映出了档案学的研究相对封闭^[37]；多数图书馆学、情报学期刊具有良好的区分能力，区分性较好；JDC 为负的期刊则没有表现出良好的差异性。此外，由于 PCA 降维会存在失真的情况，所以导致个别期刊分布出现偏差。

(2) 代表性期刊的 JDC 特征鲜明。①《数据分析与知识发现》的 JDC 在图书情报期刊中最大，说明该期刊的个性化特征非常突出，这可能与其偏向于计算机、数据科学等研究方向有关。②《情报学报》和《中国图书馆学报》作为一流期刊，其 JDC 呈现出两极分化。前者的研究结合了更多新兴主题而更具特色，这可能是其创新力所在；而后者的研究多是领域共同关注的核心问题，具有深度，这正映射了其在学科领域中的重要地位。③JDC 为负的期刊中，《图书馆学研究》和《现代情报》并非综

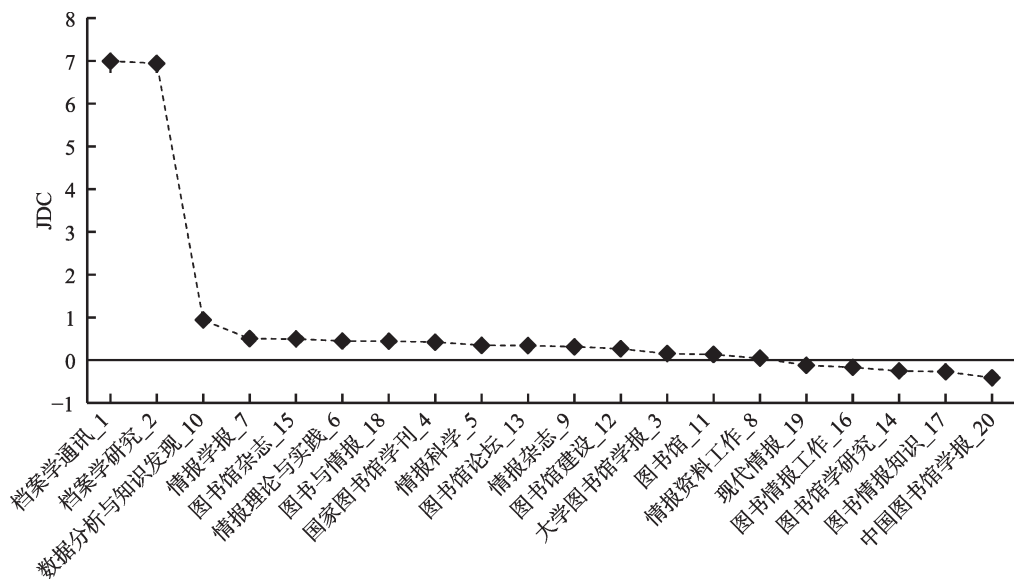


图3 LIS学科20个期刊的JDC散点图

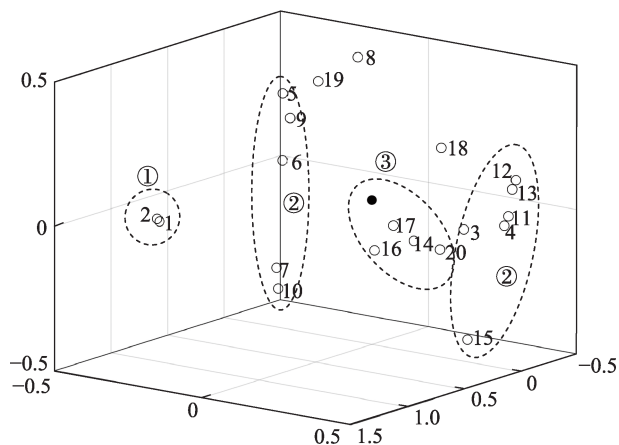


图4 2017年LIS学科20个期刊的空间分布

合类期刊,这说明它们的研究出现了一定程度的“同质化”,这对于期刊的特色发展可能是不利的。

基于上述分析,LIS学科中期刊的差异性出现了分层,个体之间的差异较为明显。档案学期刊的差异程度最大,其研究内容最具特色;《数据分析与知识发现》和《情报学报》表现出了较大的差异性,研究特色也较为突出;《中国图书馆学报》表现出了显著的综合性特征,而《图书馆学研究》和《现代情报》则表现出了“同质化”趋向。

4.1.3 不同时间跨度期刊的JDC对比分析

基于单年度数据计算所得的JDC可能具有一定的偶然性和局部性,为此,笔者采用2013—2017年LIS期刊的抽样数据来对5年度的JDC进行测算,结果如图5所示;并利用PCA对JS进行降维(图6)来辅助分析单年度和5年度期刊的差异性特征。

观察图5发现,2013—2017年JDC为正的期刊

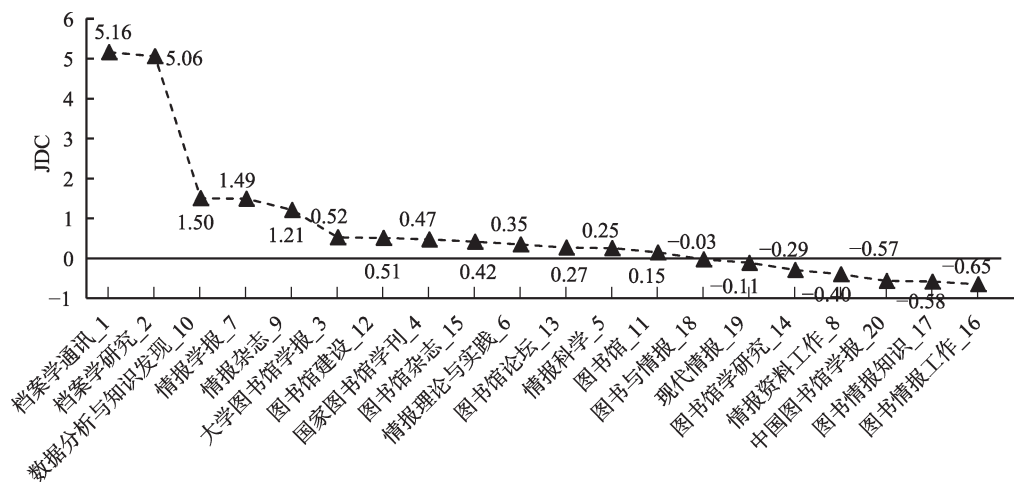


图5 2013—2017年LIS学科期刊JDC

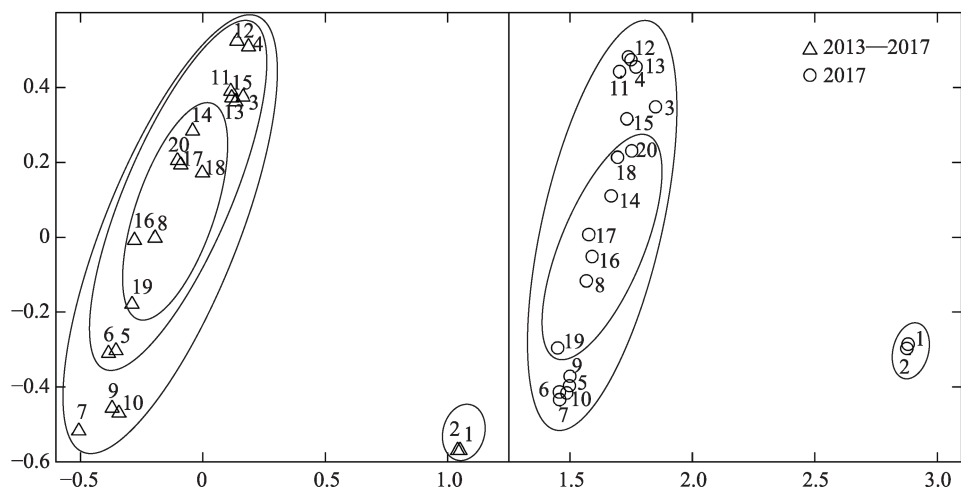


图6 单年度和5年度 LIS 期刊空间分布

个数仍占据大多数，说明大部分 LIS 期刊在 5 年时间跨度内仍保持了较好的区分性，但与单一年份的 JDC 相比也呈现出一些不同的特征。①与图 3 相比，图 5 中档案学期刊与图书情报类期刊的 JDC 差距缩小，说明较大的时间跨度使得档案学的研究特色被弱化，内容的差异程度在缩小。②图书情报类期刊的差异程度变大，JDC 分层更加明显。从图 6 可以看出，5 年度期刊分布的离散度更大，其 JDC 值的差距也被拉大了。同时，情报学的 3 个期刊 JDC 增大，与其他期刊形成分层。③期刊个体的差异性特征被充分展现。5 年中《情报学报》(7) 和《情报杂志》(9) 的 JDC 增大，其研究内容差异性得到凸显，而《情报资料工作》(8) 和《图书与情报》(18) 的 JDC 为负，表明其研究内容与其他期刊无显著差异，可能存在“同质化”趋向。

4.2 LIS 期刊的 JDC 年度趋势分析

为了探究 JDC 在时间维度上的变化规律，采用 LIS 学科 20 个期刊 2013—2017 年的抽样数据作为实验样本，分别计算每个年度的 JDC。但由于各年度 JDC 是基于不同的 JJM 计算所得，无法直接进行分析，所以笔者采用 JDC 排名来探测其变化规律，结果如表 3 所示。为了能够更加细致地分析期刊的变化轨迹，首先对 JDC 排名折线的斜率（也即后年与前年的 JDC 排名差值）进行层次聚类，结果如图 7 所示，大体上可以将期刊变化轨迹划分为 4 类，据此绘制折线图（图 8）。

（1）LIS 期刊的 JDC 排名变化表现出规律性，其变化轨迹大体呈现 4 种趋势。①图 8a 中的期刊变化趋势总体上呈扁平的“M”形，在 2015 年和 2017

表 3 2013—2017 年 LIS 期刊各年度 JDC 排名

编号	2013 年	2014 年	2015 年	2016 年	2017 年
1	2	2	1	1	1
2	1	1	2	2	2
10	3	7	5	3	3
7	4	5	7	7	4
15	9	4	3	12	5
6	7	11	11	10	6
18	14	10	8	18	7
4	10	6	9	6	8
5	6	13	12	9	9
13	17	12	10	5	10
9	5	3	4	4	11
12	13	14	6	11	12
3	8	9	13	8	13
11	12	8	15	13	14
8	16	19	16	15	15
19	11	16	14	14	16
16	19	18	19	20	17
14	18	17	17	17	18
17	20	15	18	16	19
20	15	20	20	19	20

年可能由于某些趋同性研究而出现下降趋势。②图 8b 中期刊的变化趋势比较平缓，表明这些期刊个性化发展比较稳定。③图 8c 中期刊的 JDC 排名在大幅下降之后又平缓上升或稳定发展，说明近年来这些期刊的差异性特征持续向好，稳中带升，研究特色逐渐显现。④图 8d 中期刊的 JDC 排名呈现出先升后降再升的变化趋势，在 2016 年出现较大幅度下降，其研究可能出现了“同质化”趋向。

（2）图、情、档学科期刊的 JDC 排名变化规律

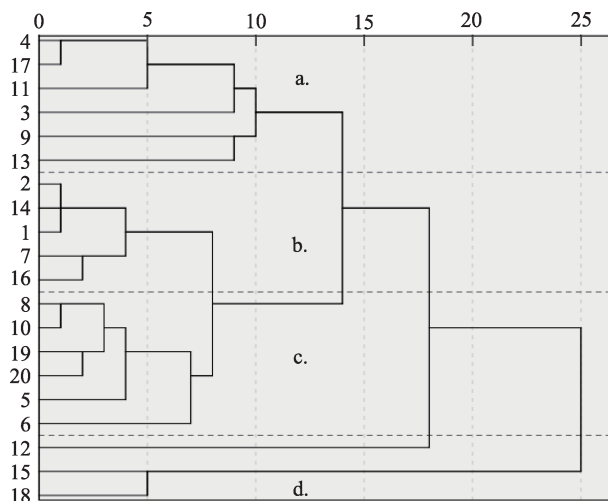


图7 基于斜率的JDC排名变化趋势聚类结果

各不相同。①档案学期刊的JDC排名一直处于前2名，说明其发展趋势非常稳定，这也反映出该学科

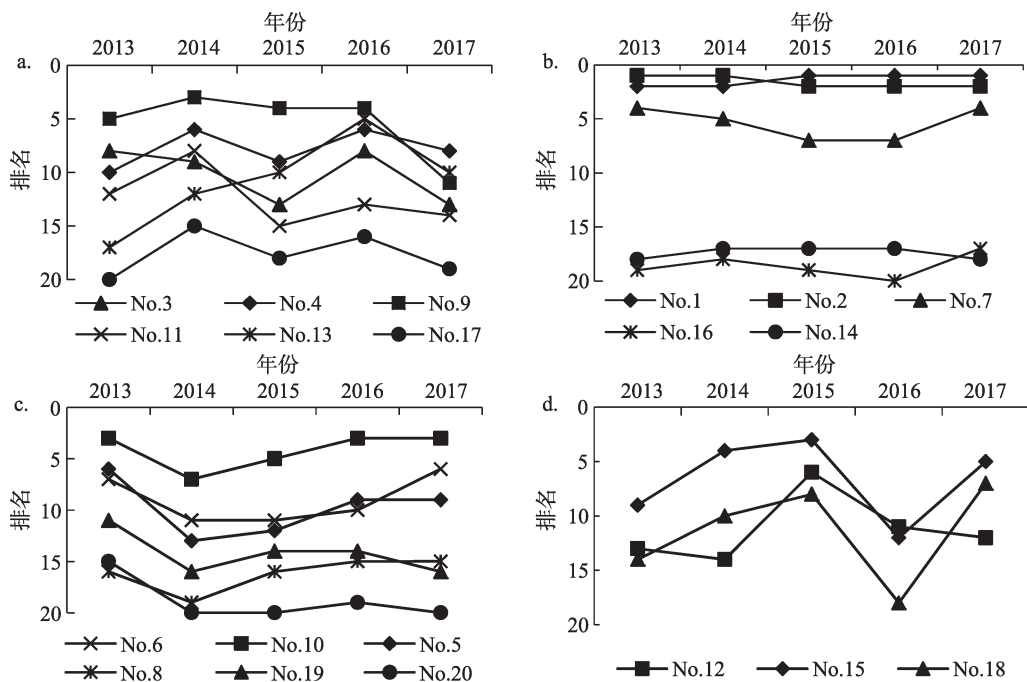


图8 LIS期刊JDC排名年度变化趋势

综上所述，LIS期刊的JDC在时间维度上具有明显的变化规律，期刊个性化或综合性特征越显著其变化趋势越稳定。档案学期刊的JDC变化最为稳定，情报学期刊的研究特色发展逐渐向好，图书馆学期刊的变化波动幅度最大，部分期刊表现出趋同性发展的趋势。

4.3 学科间JDC比较分析

为了探究学科间的期刊差异性特征，选取 LIS

研究一直相对独立，这对其学科发展和创新可能不利。②情报学期刊总体呈先降后升的发展态势，说明近几年情报学期刊的研究特色逐步凸显，个性化特征不断加强，可能处于创新发展阶段。③图书馆学期刊的变化轨迹波动较大，部分期刊在2017年出现下滑，表现出趋同性发展的态势。

(3) 不同期刊的JDC排名变化轨迹呈现出不同的特征。①JDC越大或越小的期刊变化趋势越平稳。像《情报学报》(7)和《中国图书馆学报》(20)的变化趋势比较平稳，说明个性化或综合性特征越显著的期刊，变化趋势可能就越稳定。②JDC处在中间层次的期刊排名变化波动幅度较大，反映出这些期刊研究特色发展不足。③《图书馆论坛》(13)在2017年之前都处于上升阶段，说明其在不断地提升研究特色，但之后出现下滑，可能出现了趋同性研究。

等不同门类的5个学科期刊作为研究对象，从学科维度对期刊个体及总体的JDC进行横向对比分析。

4.3.1 5个学科期刊JDC分析

采用2017年5个学科期刊的题录数据，通过随机抽样共获取4000条数据，对其进行预处理及JDC测算，结果如表4所示。

从整体上看，5个学科中有81个期刊的JDC为正，说明在该群体中大部分期刊的JDC都具有区分

表 4 2017 年 5 个学科期刊的 JDC 结果

图情档		航空航天		生物科学		艺术学		法学	
编号	JDC	编号	JDC	编号	JDC	编号	JDC	编号	JDC
3	3.56	25	0.57	51	1.74	77	1.64	100	2.25
12	3.49	36	0.52	48	1.41	76	1.51	98	2.11
4	3.40	26	0.49	53	1.31	80	1.38	83	1.98
18	3.37	31	0.48	59	1.27	79	1.09	85	1.76
14	3.31	38	0.47	52	1.27	61	0.98	81	1.61
11	3.22	32	0.46	60	1.11	74	0.86	82	1.59
13	3.20	34	0.36	50	1.03	64	0.80	96	1.42
15	2.38	29	0.26	58	0.98	63	0.69	86	1.35
19	2.25	35	0.18	43	0.90	68	0.27	89	1.34
20	2.23	23	0.17	49	0.74	70	0.18	90	1.31
16	2.03	28	0.14	57	0.67	66	0.09	94	1.30
17	1.14	40	0.13	54	0.65	73	0.03	87	1.24
8	1.06	22	-0.05	55	0.56	67	-0.06	91	1.22
5	0.87	27	-0.10	42	0.27	72	-0.20	84	1.16
7	0.84	33	-0.14	45	0.27	69	-0.21	99	1.14
6	0.68	24	-0.19	46	0.19	71	-0.38	93	1.13
9	0.64	21	-0.25	56	0.14	75	-0.45	97	0.98
2	0.58	39	-0.26	44	0.04	62	-0.52	88	0.66
1	0.56	30	-0.28	47	-0.57	78	-0.65	95	0.58
10	-0.35	37	-0.41	41	-0.66	65	-0.76	92	0.32

性。但同时，19 个 JDC 为负的期刊中有 16 个来源于艺术学和航空航天学科，为探究这种差异，采用 PCA 法将 5 个学科的 JS（100×100 维）映射至三维空间，以便观测它们在空间中的聚集状况。结果如图 9 所示，图中实心黑点为期刊空间中心。

5 个学科期刊在空间中大体分布在 3 个区域，每个学科期刊之间的差异比较明显。法学期刊分布较为集中，与其他学科期刊保持较大距离，说明该学科期刊的研究相对独立，特色鲜明，所以其 JDC 没有负值；LIS 期刊在空间中的离散度较大，个别期刊与艺术学等期刊距离较近，说明这些期刊的研究与其他学科可能存在交叉，所以有个别 JDC 也为

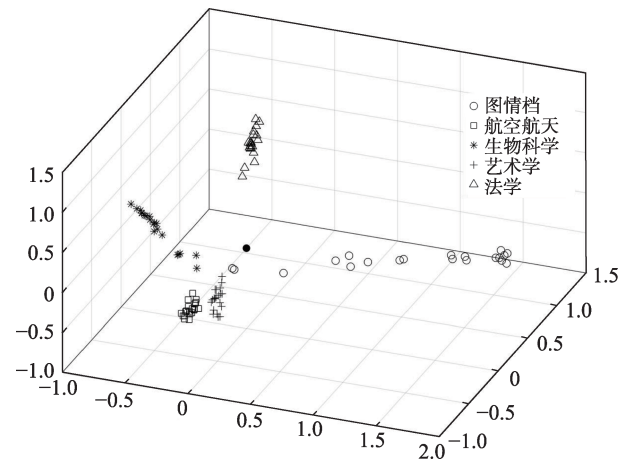


图 9 5 个学科期刊的空间分布

负值；虽然在空间分布中生物科学、航空航天和艺术学期刊的学科界限比较清晰，但它们整体处在同一区域，并且各自学科内部分布比较集中，说明这 3 个学科期刊的研究内容在某些方面具有关联性，所以导致其 JDC 普遍较小且出现较多负值。当然，由于降维存在一定的失真，只能大致观测期刊的空间分布，而不同学科期刊的差异性特征还需要进一步探讨。

为了能够更加直观地对不同学科的 JDC 特征进行深入分析，将表 4 中的数据绘制成分组直方图，如图 10 所示。

（1）5 个学科 JDC 的数值分布差异明显。①LIS 期刊 JDC 整体较大，内部出现明显分层。说明 LIS 期刊的研究较为独特、区分性较强，但也有个别期刊存在学科交叉，差异性较弱。②法学的 JDC 均为正值，说明该学科期刊研究相对独立、特色鲜明，具有较强的区分性。③生物科学的 JDC 整体处于中等水平，表明该学科期刊也具有较好的区分性。④艺术学和航空航天期刊的 JDC 均出现了两极分化，但艺术学期刊的分化更为突出，说明这两个学科期刊涉及的研究领域比较广泛，不同期刊的研究

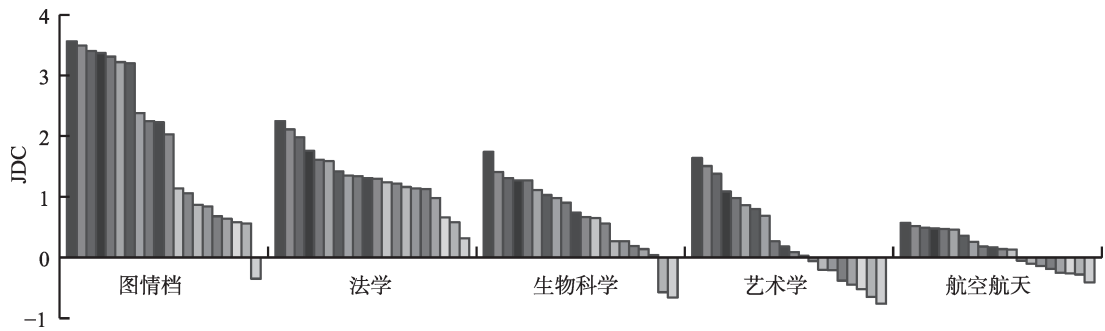


图 10 5 个学科期刊 JDC 直方图

主题存在较大差异。

(2) JDC 较大或较小的期刊在学科间分布不均衡。①JDC 较大的期刊均来自 LIS 的图书馆学, 说明图书馆学研究在 5 个学科中非常独特, 具有很好的区分性。②JDC 为负的期刊主要来自于艺术学和航空航天, 说明这两个学科的部分期刊的研究范围较大, 与其他学科具有较多的交叉领域, 如 JDC 最小的艺术学期刊《建筑学报》(65), 其研究领域横跨了多个学科。

(3) 在单学科和多学科中 LIS 期刊 JDC 表现相异。5 个学科中 LIS 期刊的图书馆学 JDC 最大, 档案学期刊 JDC 相对较小; 而在单个学科中, 档案学期刊 JDC 最大, 图书馆学期刊相对较小。说明在不同期刊群体中, 期刊的差异程度会出现相对变化。这是因为档案学作为实践学科其研究可能涉及其他学科领域, 这种交叉特征削弱了其差异性; 而图书

馆学的研究专注于图情领域, 在多学科中这种独特性就会比较显著, 从而区分性比较突出。

从上述分析可以发现, JDC 具有显著的学科特征, LIS 期刊的 JDC 整体处于较高水平, 法学期次之, 航空航天期刊最小; 学科研究领域越小, 期刊的区分性可能越大, 反之, 则可能越小; 期刊在单学科和多学科中会呈现不同的差异性特征, 这与其交叉研究的领域有关。

4.3.2 不同学科总体 JDC 分析

为了探讨不同学科总体的 JDC 特征, 笔者采用 JDC 均值来进行分析。同时采用 2017 年、2015—2017 年、2013—2017 年 5 个学科的题录数据测算期刊总体的 JDC, 从而探测不同时间跨度期刊总体的 JDC 特征。将计算结果绘制成分组直方图, 如图 11 所示。

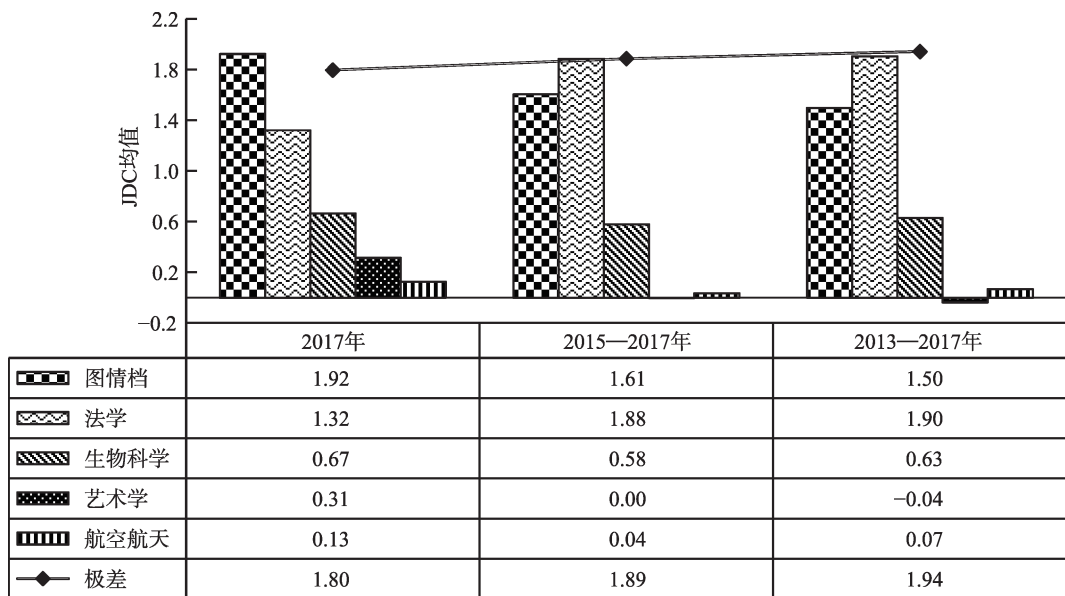


图 11 不同学科期刊总体 JDC 结果

从图 11 可以看出, 2017 年学科总体 JDC 顺序为: 图情档>法学>生物科学>艺术学>航空航天, 学科之间的区分性差异非常明显。图情档和法学属于社会学科, 与其他学科研究的交叉领域非常少, 所以表现出了较大的差异性; 而生物科学、航空航天等理工科期刊的研究内容可能存在学科交叉, 所以学科总体 JDC 就会相对较小; 艺术学属于人文科学, 但由于其研究领域的广泛性, 也存在较多交叉领域, 所以其总体 JDC 也是比较小的。

2015—2017 年和 2013—2017 年 5 个学科总体 JDC 排序均为: 法学>图情档>生物科学>航空航天>

艺术学。可以看出, 当时间跨度从 1 年扩展到 3 年时, 法学期刊研究的独特性得到充分体现, 扩大了与其他学科的差异程度; 而随时间的扩大, 艺术学研究领域的差异性被大大弱化了。当时间跨度从 3 年扩展到 5 年时, 期刊总体 JDC 却没有发生显著变化, 说明当时间跨度继续扩大时, 各个学科期刊的差异程度基本稳定。但从极差的角度看, 随时间跨度的扩大, 学科间的差异程度也会随之扩大。

5 结 论

本文基于差异性角度提出了 JDC 这一指标和测

算方法来对期刊研究内容的差异程度进行定量分析及评价。通过对 LIS 期刊差异性的测算和分析发现, 档案学期刊的差异程度最为显著, 情报学和图书馆学期刊则相对较弱;《情报学报》、《中国图书馆学报》等期刊的研究特色较为突出,《现代情报》、《图书馆学研究》等期刊则表现出了“同质化”趋向。从时间维度来看, JDC 具有明显的时间特性, LIS 学科期刊的 JDC 变化呈现出一定的规律性, 档案学期刊的 JDC 变化较为平缓, 情报学期刊差异性发展稳中向好, 而图书馆学期刊的个性化发展不太稳定。从学科维度来看, 研究领域越小, 期刊的区分性越大, LIS 和法学期刊的 JDC 普遍较大, 艺术学和航空航天普遍较小且出现两极分化; LIS 等社会科学的总体 JDC 要明显高于航空航天等理工学科, 呈现出了鲜明的学科特征。研究结果表明, 本文所提出的 JDC 方法能够有效度量学术期刊的差异性, 挖掘特色鲜明或具有“同质化”趋向的期刊个体, 能够适用于学术期刊的差异性评价。

当然, 本文也存在一些不足之处。在实验中笔者一直遵循等量原则来筛选实验对象及样本, 但实际上, 不同学科期刊数量以及期刊的有效题录量都存在较大差异, 直接采用不同规模的数据来进行实验是否会对结果造成较大影响? 在后续的研究中将就这些问题开展进一步研究, 以便验证 JDC 这一方法的普适性, 从而促进该方法在学术评价中的广泛应用。

参 考 文 献

- [1] Zhang F L. Evaluating journal impact based on weighted citations[J]. *Scientometrics*, 2017, 113(2): 1155-1169.
- [2] Salton G, Yang C S. On the specification of term values in automatic indexing[J]. *Journal of Documentation*, 1973, 29(4): 351-372.
- [3] Salton G, Yang C S, Yu C T. A theory of term importance in automatic text analysis[J]. *Journal of the American Society for Information Science*, 1975, 26(1): 33-44.
- [4] Pushpalatha K P, Raju G. Analysis of algorithms used to compute term discrimination values[C]// *Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research*. IEEE, 2010.
- [5] 张慧玲, 董坤, 许海云. 学术期刊影响力评价方法研究进展[J]. *图书情报工作*, 2018, 62(16): 132-143.
- [6] Lewis B R, Templeton G F, Luo X. A scientometric investigation into the validity of IS journal quality measures[J]. *Journal of the Association for Information Systems*, 2007, 8(12): Article No. 35.
- [7] 万昊, 谭宗颖, 朱相丽. 同行评议与文献计量在科研评价中的作用分析比较[J]. *图书情报工作*, 2017, 61(1): 134-152.
- [8] Garfield E. Long-term vs. short-term journal impact: does it matter?[J]. *Scientist*, 1998, 12(3): 11-12.
- [9] Vinkler P. Introducing the current contribution index for characterizing the recent, relevant impact of journals[J]. *Scientometrics*, 2009, 79(2): 409-420.
- [10] 俞立平. 历史影响因子: 一个新的学术期刊存量评价指标[J]. *图书情报工作*, 2015, 59(2): 89-92.
- [11] Moed H F. Measuring contextual citation impact of scientific journals[J]. *Journal of Informetrics*, 2010, 4(3): 265-277.
- [12] Braun T, Glänzel W, Schubert A. A Hirsch-type index for journal[J]. *Scientometrics*, 2006, 69: 169-173.
- [13] Egghe L. Theory and practise of the g-index[J]. *Scientometrics*, 2006, 69(1): 131-152.
- [14] Woeginger G J. An axiomatic characterization of the Hirsch-index[J]. *Mathematical Social Sciences*, 2008, 56(2): 224-232.
- [15] Kosmulski M. A new Hirsch-type index saves time and works equally well as the original h-index[J]. *ISSI Newsletter*, 2006, 2(3): 4-6.
- [16] Prathap G. Is there a place for a mock h-index?[J]. *Scientometrics*, 2010, 84(1): 153-165.
- [17] 俞立平, 李守伟. 标准特征因子的特点与应用分析[J]. *中国科技期刊研究*, 2016, 27(9): 990-993.
- [18] González-Pereira B, Guerrero-Bote V P, Moya-Anegón F. A new approach to the metric of journals' scientific prestige: The SJR indicator[J]. *Journal of Informetrics*, 2010, 4(3): 379-391.
- [19] Guerrero-Bote V P, Moya-Anegón F. A further step forward in measuring journals' scientific prestige: The SJR2 indicator[J]. *Journal of Informetrics*, 2012, 6(4): 674-688.
- [20] 李超. “HIF 指数”评价科技期刊学术影响的机理与实践[J]. *情报理论与实践*, 2011, 34(7): 44-48.
- [21] 宋晓晨, 李梦豪, 周良. 一种新型期刊评价方法——基于论文作者简介的分析[J]. *情报学报*, 2018, 37(9): 874-881.
- [22] Zhang C, Liu X, Xu Y C, et al. Quality-structure index: A new metric to measure scientific journal influence[J]. *Journal of the American Society for Information Science and Technology*, 2011, 62(4): 643-653.
- [23] 徐芳, 刘文斌. 从知识积累角度评价 SCI 期刊学术质量[J]. *情报学报*, 2013, 32(10): 1075-1089.
- [24] 马峥, 潘云涛, 武夷山. 基于引文分析的科技期刊竞争压力评价及学科间比较研究[J]. *情报学报*, 2013, 32(10): 1026-1036.
- [25] 王雯霞, 刘春丽. 不同学科间论文影响力评价指标模型的差异性研究[J]. *图书情报工作*, 2017, 61(13): 108-116.
- [26] 刘芳, 朱沙. 学术期刊与学术成果影响力主要评价指标差异性研究——以 *Nature* 期刊为例[J]. *情报杂志*, 2015, 34(8): 65-69.
- [27] 汪新红, 王国红. 学术期刊主要评价指标的学科差异性研究[J]. *科技与出版*, 2013(2): 85-88.
- [28] 盛丽娜, 顾欢. SSCI 收录期刊不同学科 Article 和 Review 参考文献量的差异性分析[J]. *中国科技期刊研究*, 2018, 29(11):

- 1153-1159.
- [29] 张丽华, 田丹, 曲建升. 中国学者发表会议论文的领域差异性与载体差异性研究[J]. 情报杂志, 2018, 37(6): 133-140.
- [30] Thelwall M. Differences between journals and years in the proportions of students, researchers and faculty registering Mendeley articles[J]. *Scientometrics*, 2018, 115(2): 717-729.
- [31] Yoon H Y, Bang H S, Woo S H. A comparative study on the logistics research between international and Korean journals[J]. *The Asian Journal of Shipping and Logistics*, 2016, 32(3): 149-156.
- [32] 柏志安, 曾剑平. 基于重叠度与完整度的 LDA 主题优选方法[J]. *计算机工程与应用*, 2019, 55(12): 155-161.
- [33] 逯万辉. 基于深度学习的学术期刊选题同质化测度方法研究[J]. *情报资料工作*, 2017(5): 105-112.
- [34] 田大芳, 张瑞丽, 魏瑞斌. 基于关键词的期刊发文的相似性测度研究[J]. *现代情报*, 2018, 38(11): 105-108, 160.
- [35] 叶鹰. 国际学术评价指标研究现状及发展综述[J]. *情报学报*, 2014, 33(2): 215-224.
- [36] Zhang J, Korfhage R R. A distance and angle similarity measure method[J]. *Journal of the American Society for Information Science*, 1999, 50(9): 772-778.
- [37] 韩毅, 李健. 图书馆学、情报学与档案学的共性与差异分析[J]. *情报资料工作*, 2012(4): 5-10.

(责任编辑 魏瑞斌)