

DOI:10.12154/j.qbzlgz.2019.06.005

# 基于 $z$ 指数的AAS高关注度学科研究主题识别\*

牌艳欣 李长玲 刘运梅 (山东理工大学科技信息研究所 淄博 255049)

**摘要** 为了准确、有效地识别学科研究热点与前沿主题,文章选择Altmetric.com平台17个指标加权综合的Altmetrics Attention Score(以下简称AAS)值,基于 $z$ 指数方法识别高关注度研究主题。选择SSCI数据库中情报学5种优秀期刊2018年的论文为样本,获取这些论文的AAS值,将AAS值替代 $z$ 指数中的被引频次,构建 $z_i$ 指数模型识别高AAS关注度研究主题,构建平均数基准线分类模型对识别结果进行细分。识别出情报学的16个高关注度研究主题,并细分潜力类、突现类、核心类、边缘类4类学科研究内容,发现本文构建的 $z_i$ 指数模型和高关注度学科研究主题分类模型是可行、有效的。

**关键词**  $z$ 指数  $z_i$ 指数 Altmetrics Attention Score 指标 高关注度研究主题

## AAS High Concern Subject Research Topic Identification Based on $z$ -index

Pai Yanxin Li Changling Liu Yunmei

(Institute of Science and Technical Information, Shandong University of Technology, Zibo, 255049)

**Abstract** In order to identify subject research hotspots and cutting-edge topics accurately and effectively, choosing the Altmetrics Attention Score(AAS) value of the Altmetric.com platform, which is weighted synthesis of 17 indicators, to identify high-profile research topics based on the  $z$ -index. Articles of five excellent information science journals in 2018 selected from SSCI database as samples to obtain the AAS values of these papers. The citation frequencies in the  $z$ -index are replaced by AAS, and a  $z_i$  index model was constructed to identify the research topics. The authors construct  $z_i$  index model to identify high AAS attention research topics, and construct a mean baseline classification model to subdivide the recognition results. There are 16 high-interest research topics of information science were identified, and the research contents of four categories of potential, emerged, core, and marginal categories were subdivided. Research found that the  $z_i$  index model and the topic classification model of high-concerned subject research constructed in this paper are feasible and effective.

**Keywords**  $z$  index,  $z_i$  index, Altmetrics Attention Score, highly concerned subject topic

## 1 引言

科学的本质是知识的创造性活动。大量研究指出获取足够关注度的研究主题,极具潜力发展成为科学研究中实现突破性创造的关键研究领域。同时,对学

科高关注度研究主题识别与分析,有助于了解学科发展现状与趋势,把握最新的研究方向。因此,高关注度研究主题的识别是科研工作者长期关注的焦点。

科学文献作为科学研究成果的主要载体,是科学研究参考、借鉴的主要依据。所以,国内外许多学者以

\*本文系国家自然科学基金重点项目“跨学科潜在知识生长点识别与创新趋势预测研究”(编号:19ATQ006)的研究成果之一。

科学文献为数据来源,采用文献计量分析方法识别学科研究主题。Glänzel等<sup>[1]</sup>采用文献耦合分析、文本挖掘、主题聚类等方法识别新兴领域的热点研究主题。Tu等<sup>[2]</sup>提出新颖度指数,基于文献计量与文本内容分析相结合的方法识别新兴主题。国内学者黄鲁成等<sup>[3]</sup>通过提取科学文献摘要关键词,构建多属性测度模型识别新兴主题。次年通过引文分析,利用SAO结构算法、语义相似度算法、多维尺度分析法等识别新兴主题<sup>[4]</sup>。周鑫等<sup>[5-6]</sup>通过关键词词频、全文分词的有效语词词频分析,识别美国情报学与我国情报学的学科研究热点。

近年来,伴随在线学术资源平台、网络社交媒体的成熟发展,用户学术交流模式发生极大改变,网络媒体数据呈指数型增长,国内外学者积极采用Altmetrics方法识别学科热点主题。Takahashi等<sup>[7]</sup>以作者博客中的关键词为研究对象,筛选突现词汇,确定当前的热点主题。盛宇<sup>[8]</sup>研究新浪微博的文本数据,提出一种学科热点发现、跟踪和分析机制,挖掘热点主题。赖纪瑶等<sup>[9]</sup>选择知乎网站“情报学”话题“精华页面”的回答文贴,挖掘用户群高关注的学科焦点主题。王贤文等<sup>[10]</sup>基于科学文献被下载的即时信息,实时追踪某一领域的研究趋势、挖掘研究热点、探测研究前沿。赵雅馨等<sup>[11]</sup>利用Altmetrics方法中7项指标测度科学文献的综合关注度,建立数据筛选与过滤机制,挖掘领域的研究热点及前沿主题。段庆锋等<sup>[12]</sup>利用社交媒体数据,构建主题关注热度与强度监测指标,识别高关注度增长潜力的学科新兴主题。

科学文献具有权威性高、可信度强等不可替代的优势。但传统计量分析方法识别的学科前沿主题难以反映“学术圈群体”以外更广泛的社会大众对于该学科的关注热点和知识需要。目前用Altmetrics方法对网络媒体数据进行分析,识别学科热点的相关研究局限于一种或几种指标数据的分析,识别结果不够全面。

Altmetrics Attention Score指标是较为成熟、被广泛认可的一种Altmetrics计量算法。它是Altmetric.com综合平台上全部Altmetrics指标数据,包括新闻报道数、博客帖子数、Mendeley读者数、同行评审数等17个在线政策文件与主流媒体、在线社交媒体、在线参考文献管理工具、其他在线来源的使用、下载、推广等数据,通过一定的算法加权得到的一个聚合指标<sup>[13]</sup>。2017年,Altmetric.com网站将Web of Science中的被引频次

纳入其指标体系。因此AAS值是科学文献传统评价与网络媒体评价的综合反映,可以更加全面地评估科学文献的学术水平,以及发表后所受到的在线关注度。

近几年,z指数方法在p指数的基础上增加了一致性指标,有效地协调了被引频次与发文数量的关系,在论文、人才、机构、期刊等评价方面表现出较好的评价效果。因此本文尝试选择科学文献的网络媒体表现数据——AAS值,基于z指数方法进行高关注度学科研究主题的识别,以更综合、全面掌握学科前沿动向。

## 2 z指数及其在AAS高关注度学科研究主题识别中的应用

### 2.1 z指数

z指数是2014年由印度学者Prathap<sup>[14]</sup>在 $p = \left[ c \times \frac{c}{n} \right]^{\frac{1}{3}}$ 指数的基础上,引入描述被引分布一致性指标 $v$ 而得到的新型复合计量指标,其本质是将表示数量、质量、一致性程度的三个指标等权重相乘,故也称其为Quantity(数量)-Quality(质量)-Consistence(一致性)的3D效能评价指标。其计算公式如下:

$$z = [pv]^{\frac{1}{3}} = \left[ c \times \frac{c}{n} \times \frac{p}{\sum_{k=1}^n ck^2} \right]^{\frac{1}{3}} = \left[ c \times \frac{c}{n} \times \frac{\frac{c^2}{n}}{\sum_{k=1}^n ck^2} \right]^{\frac{1}{3}} = \left[ \frac{\frac{c^4}{n^2}}{\sum_{k=1}^n ck^2} \right]^{\frac{1}{3}} \quad (1)$$

其中 $c$ 表示论文的总被引频次,表征数量; $n$ 表示文章数量, $c/n$ 篇均被引量表示质量; $c_k$ 表示第 $k$ 篇文章的被引频次, $k=1,2,\dots,n$ , $v = \frac{p}{\sum_{k=1}^n ck^2}$ 一致性指标,在 $p$ 指数一定的情况下,一篇或几篇论文的高被引无疑会使平方和变大, $v$ 变小。但无论是人才、团队、机构、学术期刊,在整体水平较高的情况下,个体差距越小越好,即 $v$ 值越大越好。因此,一致性指标 $v$ 的引入有效解决了特异值问题。

随后,Prathap<sup>[15]</sup>对 $z$ 指数中的指标内涵与特征做了进一步的解释说明,分析了 $z$ 指数与 $h$ 指数之间的关系,阐述了用 $z$ 指数进行科研水平影响力评价<sup>[16]</sup>和最新研究成果评价<sup>[17]</sup>的有效性和可操作性。

国内对 $z$ 指数的研究刚刚起步,何晓庆等<sup>[18-20]</sup>通过

分析 $z$ 指数的内涵及计算方法,论证了 $z$ 指数评价学者、期刊、科研机构的合理性与有效性,认为 $z$ 指数是一个比 $h$ 指数、 $p$ 指数更具备整体性、全面性的综合评价指标,更具高区分度、高灵敏度等特性。俞立平等<sup>[21]</sup>分析了 $z$ 指数与相关文献计量指标之间的关系,针对 $z$ 指数存在的缺陷,提出改进指数 $z_n$ ,用于期刊评价。

目前,国内外学者利用 $z$ 指数进行评价的主体主要涉及国家、科研机构、期刊、学术论文、学者等,尚未有利用 $z$ 指数进行主题识别方面的相关研究。本文利用 $z$ 指数方法,基于网络社交媒体中的AAS数据,构建一个测度文献主题受关注程度大小的 $z_i$ 指数,识别学科高关注度研究主题。以情报学领域的科学文献为例,论证利用 $z_i$ 指数方法识别高关注度学科研究主题的可性。

## 2.2 $z$ 指数应用于学科AAS高关注度研究主题识别

学术论文作为科学文献最基本也是最重要的表现形式,其使用、下载、引用等指标值在一定程度上反映了其研究主题的被关注程度,基本思想如图1所示。

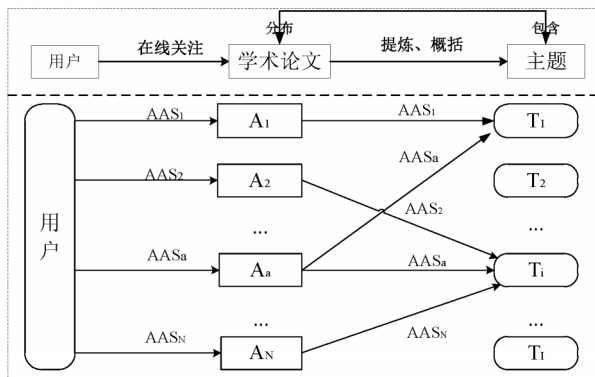


图1 利用AAS值测度主题关注度原理图

如图1所示,一篇学术论文可包含多个研究主题,一个主题亦可分布于多篇论文中。因此,研究主题和学术论文之间存在多对多的关系。用户通过阅读、下载、推荐、讨论等不同类型的学术行为,表现出对学术论文不同程度的关注,实际亦是对学术论文所研究主题的同等级度的关注<sup>[21]</sup>。本文将Altmetrics中评价单篇论文在线关注度大小的参数指标AAS引申为测度研究主题在线关注度大小的指标,以识别学科前沿热点。图1中,主题 $T_i$ 分布于 $A_2$ 、 $A_a$ 、 $A_n$ 共3篇学术论文中,分别获取3篇论文的关注度指数 $AAS_2$ 、 $AAS_a$ 、 $AAS_n$ ,则主题 $T_i$ 的总关注度得分为三者相加,即: $AAS_2+AAS_a+AAS_n$ 。

基于 $z$ 指数方法,本文提出识别学术论文AAS高关注度研究主题的 $z_i$ 指数,其计算公式为:

$$z_i = \left[ S(T_i) \times \frac{S(T_i)}{n_i} \times C(T_i) \right] = \left[ S(T_i) \times \frac{S(T_i)}{n_i} \times \frac{\frac{S(T_i)^2}{n_i}}{\sum_{j=1}^{n_i} [S_j(T_i)]^2} \right]^{\frac{1}{3}}$$

$$= \left\{ \frac{[S(T_i)]^4}{n_i^2 \times \sum_{j=1}^{n_i} [S_j(T_i)]^2} \right\}^{\frac{1}{3}} \quad (2)$$

若某一主题 $T_i$ 共分布于 $n_i$ 篇学术论文中,则:

$S_j(T_i)$ 是主题 $T_i$ 于第 $j$ 篇论文中的单篇AAS得分量, $j=1,2,\dots,n_i$ ;  $S(T_i) = \sum_{j=1}^{n_i} S_j(T_i)$ 是研究主题 $T_i$ 的AAS得分总量,描述主题 $T_i$ 的关注度数量整体规模。

$\frac{S(T_i)}{n_i}$ 为篇均得分量,反映主题 $T_i$ 在单篇论文层面的平均关注水平,是描述主题 $T_i$ 关注度质量的指标。

$C(T_i)$ 分布一致性指标,用来描述主题 $T_i$ 在其分布的论文中受关注分布一致性情况,且 $C(T_i) \in [0, 1]$ ,作为调节因子解决极值问题。例,如果某主题 $T_i$ 只在一篇或几篇文章中的在线关注度得分明显偏高,而在其他文章关注度得分少甚至都为0,会使 $S(T_i)$ 总得分和 $\frac{S(T_i)}{n_i}$ 篇均得分都处于较高水平。引入一致性指标 $C(T_i)$ ,使得单一增加主题 $T_i$ 分布篇数,或者只在极少量论文中有高分的情况下,都不能直接提高主题 $T_i$ 的关注度,只有 $T_i$ 在所分布的论文中取得较高且较均匀的关注时,才能获得高 $z_i$ 值。这样, $z_i$ 指数从整体规模-平均水平-分布一致性三个层面描述主题 $T_i$ 的关注程度,可以综合、有效地识别学科前沿热点主题。

## 2.3 基于平均数基准线的学科AAS高关注度研究主题分类

$z_i$ 指数值较大的识别结果都是在线关注度高的学科研究主题,这些研究主题根据不同的规则条件,可以判定其目前的研究现状以及对学科的贡献程度。本文设计以下规则,将AAS高关注度研究主题细分。

首先,对高关注度研究主题识别结果,计算每个研究主题 $T_i$ 对应的论文分布篇数 $n_i$ 、AAS得分总量 $S(T_i)$ 、篇均得分 $\frac{S(T_i)}{n_i}$ 、分布一致性指标 $C(T_i)$ 的平均值,作为各指标数值的基准值。

其次,依据图2所示规则,将AAS高关注度研究主

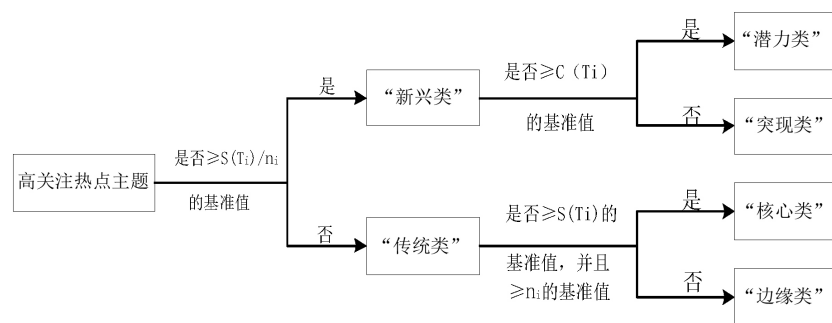


图2 AAS高关注度学科研究主题分类模型

题细分为潜力类、突现类、核心类、边缘类四类。

### 3 实证研究——以情报学为例

#### 3.1 数据来源

Web of Science数据库载文资源收录广泛,并能代表国际研究前沿。根据ISI Web of Knowledge平台的JCR报告,参考其他学者的研究经验<sup>[23-24]</sup>,本文选择能够代表情报学领域研究前沿和热点的5种核心影响力期刊作为样本来源,分别是: *Journal of the Association for Information Science and Technology*、*Scientometrics*、*Information Processing & Management*、*Journal of Documentation*、*Journal of Informetrics*。网络社交媒体中的测度指标具备实时性动态监测的优势<sup>[11]</sup>,一篇文献的受关注度大小在论文发表初期便能得到显著表现<sup>[25]</sup>,因此年份区间限定为2018年,以有效识别学科前沿主题。以“出版物名称”为检索途径,选择文献类型为“Article”,共检索到5种期刊的学术论文656篇,检索时间为2019年1月12日。

下载656篇文献的题录信息,利用Loet工具抽取每一条文献记录中DOI(文献标识码)、TI(标题)、DE(作者给定关键词)、ID(数据库标注关键词)等字段的元数据信息,构成基础数据集。在Altmetric.com网站,通过DOI号,利用Python爬虫工具,获取每篇论文的AAS值,构成本研究的实证数据集。

### 3.2 数据处理

#### 3.2.1 获取论文研究主题数据集

关键词是对文献核心内容的高度概括,因此本文选择规范化处理的关键词表征文献研究主题。过程如下:

(1)关键词数据集来源。选择作者给定关键词DE和WoS数据库标注关键词ID字段共同作为学术论文的关键词来源。这部分数据除

作者给定的关键词之外,还包括WoS数据库抽取文献标题、摘要、参考文献、引证文献等信息自动标引的关键词。这样能更全面、准确地概括论文研究主题,有效减少因作者主观因素造成的片面影响。利用Loet工具抽取实证数据集中每条记录的DE、ID字段数据,保存在Excel工作表里。

(2)关键词规范化处理。对DE、ID关键词数据集利用Wordnet词典工具,进行词干化处理,去除时态、单复数等词缀,将含义相近的词进行同义词合并,以达到描述的一致性。如处理中将collaboration in science去词干化、同义词合并处理得到scientific collaboration; content analyzing去词缀处理规范为content analysis; broader impact去比较级统一为impact;含义相近或同义词合并,如“peer review”与“referee”都表示“同行评议”,将后者规范为“peer review”;对Altmetric Attention Score、Altmetrics Score与Altmetric Aggregate Score这类对同一事物的不同描述性的新兴词汇,在掌握研究领域相关知识的基础上,结合文献调研<sup>[26-27]</sup>、访谈专家等进行人工合并处理,统一合并为Altmetric Attention Score。

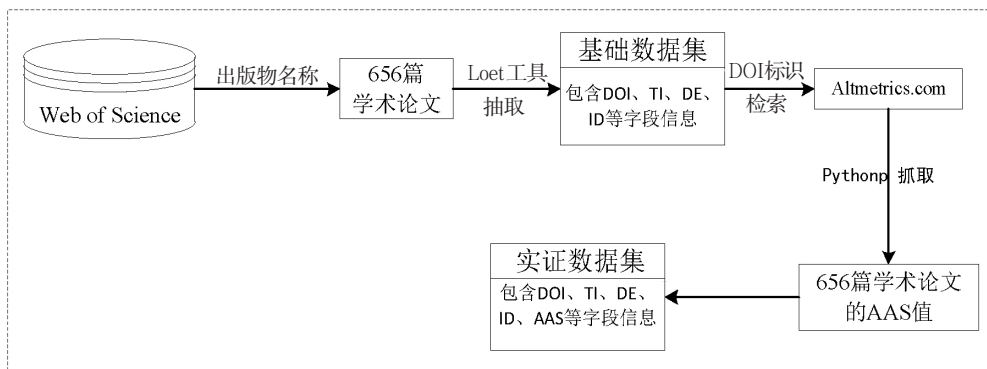


图3 数据来源流程图

(3)确定论文研究主题。经过上述处理,得到规范化的关键词487个,代表论文研究主题。样本文献的相关数据信息见表1。其中,文献字段信息IT、DE、ID见表1-3列,处理后的规范化关键词见表第4列,AAS值见表第5列。由于篇幅限制,表1仅展示其中5篇学术论文的相关信息。

### 3.2.2 计算 $z_i$ 指数值

利用Java语言编写程序,读取表中信息,统计主题 $T_i$ 的分布篇数 $n_i$ 、单篇AAS得分量 $S(T_i)$ ,计算总得分 $S(T_i)$ 、篇均得分 $\frac{S(T_i)}{n_i}$ 、分布一致性指标 $C(T_i)$ ,根据公式(2)分别计算每个主题 $T_i$ 的 $z_i$ 指数值。因为有些主题所在论

表1 2018年情报学领域5种核心期刊学术论文相关数据(部分)

题名IT	作者关键词DE	WoS关键词ID	规范化关键词	AAS
.....	.....	.....	.....	.....
Journal peer review: a bar or bridge? An analysis of a paper's revision history and turnaround time, and the effect on citation	turnaround time; peer review journal author	quality referees publication decisions authors editors fate bias	peer review turnaround time journal author publication quality editor strategy	2
Assessing the effect of the United States' "citation advantage" on other countries' scientific impact as measured in the Web of Science (WoS) database	evaluation database biased indicators bibliometrics rankings	Social-science World science collaboration performance nations system China productivity cooperation humanities	scientific impact research evaluation collaboration productivity database bibliometrics performance Social-science World science Humanity(encee) nations system China	56
Topic based research competitiveness evaluation	research competitiveness; topic model evaluation	emerging trends	research competitiveness topic model evaluation method emerging trend	5
Researchers' risk-smoothing publication strategies: Is productivity the enemy of impact?	peer review cooperation Game theory scientist strategies agent-based	science cooperation competition judgments	peer review collaboration Game theory scientist strategy agent-based competition research evaluation science	7
Reliability and accuracy of altmetric providers: a comparison among Altmetric.com, PlumX and Crossref Event Data	altmetric.com PlumX Crossref Event Data altmetrics data providers	article-level social media impact citations Altmetric Attention Score	altmetric.com PlumX Crossref Event Data altmetrics data providers article-level social media impact citations	22
.....	.....	.....	.....	.....

文的AAS值为0,所以 $S(T_i)$ 为0, $z_i$ 值等于0。 $z_i$ 值不等于0的研究主题328个,因为论文篇幅有限,取前5%,16个主题词按 $z_i$ 值由高到低排列的数据结果见表2。

表2 2018年情报学领域主题词各指标量数据结果(部分)

序号	主题 $T_i$	$n_i$		$S(T_i)$		$\frac{S(T_i)}{n_i}$		$C(T_i)$		$z_i$ 指数	
		数值	排名	数值	排名	数值	排名	数值	排名	数值	排名
1	altmetrics	36	5	456	4	12.667	5	0.324	5	12.273	1
2	social media	41	4	460	3	11.220	7	0.353	3	12.179	2
3	bibliometrics	85	1	739	1	8.694	10	0.215	14	11.221	3
4	collaboration	23	10	235	12	14.130	3	0.405	1	11.083	4
5	citation analysis	51	2	498	2	9.765	9	0.273	11	10.952	5
6	user behavior	18	12	262	5	14.556	2	0.323	6	10.687	6
7	open access	19	11	251	10	13.210	4	0.315	7	10.199	7
8	university rankings	29	8	276	7	9.517	9	0.371	2	9.906	8
9	peer review	17	13	285	6	16.764	1	0.199	16	9.849	9
10	gender	16	15	202	13	12.625	6	0.268	12	8.832	10
11	information seeking	24	9	267	9	11.125	7	0.222	13	8.679	11
12	impact	44	3	271	8	6.160	15	0.315	7	8.114	12
13	machine learning	36	5	298	5	8.280	11	0.200	15	7.902	13
14	research evaluation	34	7	237	11	6.971	14	0.277	10	7.734	14
15	research productivity	24	9	178	14	7.417	13	0.289	9	7.262	15
16	Google scholar	17	13	137	15	8.083	12	0.328	4	7.149	16
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....

### 3.2.3 AAS高关注度学科研究主题类别细分

根据表2结果,对 $z_i$ 指数值排名前16位的2018年度情报学AAS高关注度研究主题,用各指标 $n_i$ 、 $\frac{S(T_i)}{n_i}$ 、 $C(T_i)$ 的平均数作为基准线,依据图2的分类模型,运行自编程序,分类结果见表3。

表3 2018年情报学AAS高关注度研究主题分类表

主题类型	主题细分	研究主题
“新兴类”	“潜力类”	Altmetrics, social media, collaboration, user behavior, open access
	“突现类”	peer review, gender, information seeking
“传统类”	“核心类”	Bibliometrics, citation analysis, impact factors
	“边缘类”	machine learning, research evaluation, research productivity, Google scholar

## 3.3 结果分析

### 3.3.1 $z_i$ 指数可以有效识别不同类型的AAS高关注度学科研究主题

将AAS值替代 $z_i$ 指数中的被引词频,构建 $z_i$ 指数,识别高关注度研究主题。同时,基于平均数基准线构建AAS高关注度研究主题分类模型,将情报学高关注度研究主题细分为表3中的四类:

“潜力类”主题包括 Altmetrics, social media, collab-

oration、user behavior、open access,这类研究主题篇均关注度  $\frac{S(T_i)}{n_i}$  高,而且关注度的分布一致性  $C(T_i)$  也高,每篇相关文献都有较高的关注度。所以,这类研究主题具有较好的研究潜力,可能是学科研究前沿或研究热点。

“突现类”研究主题包括 peer review、gender、information seeking,这类研究主题篇均关注度  $\frac{S(T_i)}{n_i}$  高,但关注度分布一致性指标  $C(T_i)$  低,说明该类研究主题关注度高度集中在一篇或几篇相关文献上,不是每一篇论文都被高度关注。因此,这类研究主题可能是最新出现或新视角的学科研究内容。

“核心类”主题包括 bibliometrics、citation analysis、impact factors,它们的论文分布篇数  $n_i$  和关注度总得分  $S(T_i)$  都高,但由于相关研究论文的数量多,使篇均关注度  $\frac{S(T_i)}{n_i}$  低。高数量的研究论文,表明这类研究主题一直是情报学研究中最稳固的核心研究内容<sup>[28]</sup>。

“边缘类”主题包括 machine learning、research evaluation、research productivity、Google scholar,这类研究主题论文分布篇数  $n_i$ 、关注度总得分  $S(T_i)$ 、篇均关注度  $\frac{S(T_i)}{n_i}$  都低,说明近期这类研究主题的关注度不高,有下降的趋势,学科研究热度可能会持续降低。

### 3.3.2 基于 AAS 数据的 $z_i$ 指数识别高关注度“新兴类”前沿主题更有效

主题分布频次的大小在一定程度上代表着相关研究的关注程度,因此传统方法主要用词频分析、中心性分析等方法,识别高关注度的学科主题。近年,许多文献通过对情报学领域学术文献的高频词运用传统方法,识别得到高关注度热点主题有 bibliometrics、citation analysis、impact factors、scholarly communication、Google scholar 等,这些结果与本文识别的“传统类”研究主题极其相似。一方面说明本文识别方法与结果的有效性,另一方面说明传统方法无法有效识别更具潜力的“新兴类”研究主题。本文利用  $z_i$  指数不仅可以识别传统类研究主题,还能识别如 Altmetrics、social media、user behavior、open access、information seeking 等代表未来学科发展方向的高关注度新兴研究主题,该结果与被认为代表领域最新动态风向标的 ASIS&T2018 年会上多篇论文的研究结论一致<sup>[29-33]</sup>,也与段庆锋等<sup>[12]</sup>

的研究结论 scholarly publishing、open access、altmetrics 一致,或许更全面一些。因此,用 AAS 值替代被引频次的  $z_i$  指数方法识别学科“新兴类”前沿主题更有效、全面。

### 3.3.3 $z_i$ 指数较好地实现了研究主题 AAS 关注度得分总体规模、平均水平以及分布一致性之间的平衡

观察表 2 可以发现,研究主题 altmetrics 的  $z_i$  指数值最高。其分布篇数  $n_i$ 、AAS 总得分  $S(T_i)$ 、篇均得分  $\frac{S(T_i)}{n_i}$ 、分布一致性指标  $C(T_i)$  排名都比较靠前,分别位居第 5 位、第 4 位、第 5 位、第 5 位,且得分分布一致性指标  $C(T_i)=0.324$ ,说明其关注度得分在  $n_i=36$  篇论文中的分布较为均匀,36 篇 Altmetrics 研究论文都有较好的关注度,这说明 Altmetrics 是目前情报学最高关注度的研究主题。近期,Bornmann 等<sup>[34-37]</sup>都在最新研究中指出:在线科研行为活跃度的不断增强,使得以 Altmetrics 为代表的新型信息计量理论与工具成为情报学及有关领域的新兴研究内容,与该主题相关的研究保持持续的研究关注度,成为国际情报学领域的前沿研究主题。

### 3.3.4 $z_i$ 指数有效限制了 AAS 特异值的突出作用

表 2 中, $z_i$  指数排名第 9 位的研究主题 peer review,虽然篇均得分  $\frac{S(T_i)}{n_i}$  最高 16.764,但其分布一致性指数  $C(T_i)$  最小,为 0.199。在 Altmetric.com 网站,检索主题 peer review 分布的 17 篇论文的单篇关注度得分 AAS 值,其中两篇论文的 AAS 值为 82 和 66,占总得分  $S(T_i)=285$  的一半,而其中的 8 篇论文 AAS 值为 0,因此关注度分布一致性差。这说明将一致性指标  $C(T_i)$  作为调节因子引入  $z_i$  指数,可以在一定程度上有效避免因一篇或几篇文章突出的高关注度得分,造成研究主题识别结果的误差。

### 3.3.5 $z_i$ 指数能够有效限制分布篇数 $n_i$ 对研究主题 $T_i$ 关注度的影响

通过表 2 发现,按照主题分布篇数  $n_i$  和关注度总得分  $S(T_i)$  的排名,主题 bibliometrics 最高, citation analysis 次之,分别为第 1、2 位。但两者篇均得分  $\frac{S(T_i)}{n_i}$ ,由于分布篇数的数量大,分别降到了第 10 和第 9 位。这主要源于主题 bibliometrics 与 citation analysis 一直是情报学计量研究领域经久不衰的传统研究热点,研究文献数量较多,因此获得较高水平的得分总量,但相比较于

altmetrics 这类近年来新兴热点主题,篇均关注水平较低。虽然篇均关注度不高,但论文数量和关注度得分总量高,bibliometrics 与 citation analysis 的 $z_i$ 指数排名依然排名第3和第5。因此, $z_i$ 指数有效地协调了研究主题关注度的数量和质量,限制了分布篇数 $n_i$ 对主题关注度的影响。

#### 4 结语

科学文献与网络博文的词频分析,以及 Altmetrics 单指标或几个指标的分析,都无法准确地对科学文献研究主题的关注度做出实时、全面的评价。而 Altmetrics Attention Score(AAS)是对包含被引数据在内的17种不同类型的网络媒体在线实时数据,进行特定加权聚合得到的一个度量指标,从而可以多元化地及时测度学术成果的社会和学术影响力,使得评价更为全面、具体和客观。另外, $z$ 指数综合反映学者、期刊、科研机构的被引数量、被引质量、被引数量与质量一致性3D效能。本文将AAS值替代 $z$ 指数中的被引频次,构建识别AAS高关注度研究主题的 $z_i$ 指数模型,从关注数量、关注质量、关注数量与质量一致性3个方面,综合评价研究主题网络媒体表现、构建基于平均数基准线的AAS高关注度研究主题分类模型,对识别结果细分为:潜力类、突现类、核心类、边缘类4种类型,进一步发现科学研究前沿与热点。以情报学领域的科学文献为例进行实证研究,实证结果说明了该方法的可行性与有效性。

网络环境下,网络媒体数据动态性强,且不同学科的数据特征不同。因此,本文后续研究将考虑时间因素的影响,分析不同学科的高关注度研究主题识别差异,帮助不同学科的研究人员把握学科前沿与热点,推动学科与科学的进步。

#### 参考文献

- [1] Glänzel W, Thijs B. Using 'core documents' for detecting and labelling new emerging topics[J]. *Scientometrics*, 2012,91(2): 399-416.
- [2] Tu Y, Seng J. Indices of novelty for emerging topic detection[J]. *Information Processing & Management*, 2012,48(2):303-325.
- [3] 黄鲁成,唐月强,吴菲菲,等.基于文献多属性测度的新兴主题识别方法研究[J]. *科学学与科学技术管理*, 2015(2):34-43.
- [4] 黄鲁成,张璐,吴菲菲,等.基于突现文献和SAO相似度的新兴主题识别研究[J]. *科学学研究*, 2016(6):814-821.
- [5] 周鑫,蒋勋,陈媛媛.词频变化率模型视域下美国情报学研究发展动向分析[J]. *情报科学*, 2017(4):166-172.
- [6] 刘小慧,李长玲,刘运梅,等.基于作者—核心关键词2—模网络的潜在跨学科合作组合识别——以图书情报学与计算机科学为例[J]. *情报理论与实践*, 2018,41(2):105-110.
- [7] Takahashi T, Tomioka R. Discovering emerging topics in social streams via link anomaly detection[J]. *IEEE Transactions on Knowledge & Data Engineering*, 2011, 26(1):120-130.
- [8] 盛宇.基于微博的学科热点发现、追踪与分析——以数据挖掘领域为例[J]. *图书情报工作*, 2012,56(8):32-37.
- [9] 赖纪瑶,魏思仪,秦玥.基于“知乎”关注热点的学科认知特点挖掘——以情报学为例[J]. *图书情报工作*, 2017(24):96-104.
- [10] 王贤文,毛文莉,王治.基于论文下载数据的科研新趋势实时探测与追踪[J]. *科学学与科学技术管理*, 2014(4):3-9.
- [11] 赵雅馨,杨志萍.研究热点探测的替代计量学方法和应用——以信息与计算科学为例[J]. *情报杂志*, 2016(11):39-44.
- [12] 段庆峰,潘小换.利用社交媒体识别学科新兴主题研究[J]. *情报学报*, 2017,36(12):1216-1223.
- [13] How is the Altmetric Attention Score calculated?[EB/OL]. [2019-01-02]. <https://help.altmetric.com/support/solutions/articles/6000060969-how-is-the-altmetric-score-calculated/>.
- [14] Prathap G. A three-class, three-dimensional bibliometric performance indicator[J]. *Journal of the Association for Information Science and Technology*, 2014,65(7):1506-1508.
- [15] Prathap G. The synergy-index and the formula for the h-index[J]. *Journal of the Association for Information Science and Technology*, 2014,65(2):426-427.
- [16] Prathap G. A three-dimensional bibliometric evaluation of research in polymer solar cells[J]. *Scientometrics*, 2014,101(1): 889-898.
- [17] Prathap G. A three-dimensional bibliometric evaluation of recent research in India[J]. *Scientometrics*, 2017,110(3):1085-1097.
- [18] 何晓庆,王圣洁,胡琳. $z$ 指数在学者学术影响力评价中的应用[J]. *情报理论与实践*, 2018,41(5):50-54.
- [19] 何晓庆,王圣洁,胡琳. $z$ 指数在期刊评价实践中的应用研究[J]. *中国科技期刊研究*, 2018,29(5):509-514.
- [20] 何晓庆,王圣洁,胡琳.基于 $z$ 指数的科研机构评价的有效性实证研究[J]. *现代情报*, 2018, 38(05):82-86.
- [21] 俞立平,王作功. $z$ 指数评价学术期刊的适用性及其改进研究[J]. *情报学报*, 2018, 37(11): 1132-1139.
- [22] 王睿,胡文静,郭玮.高Altmetrics指标科技论文学术影响力研究[J]. *图书情报工作*, 2014(21): 92-98.
- [23] 赵蓉英,马丽娜.国际情报学核心期刊与研究热点的可视化

- 分析[J]. 情报科学, 2011, 29(8): 1238-1243.
- [24] 刘自强, 岳丽欣, 王效岳, 等. 主题演化视角下的国际情报学研究热点与前沿分析[J]. 图书馆, 2017(3):14-22.
- [25] 郝若扬. 高 Altmetrics 指标论文的特征分析及影响力分析[J]. 图书情报工作, 2018(8):107-114.
- [26] 陈小清, 刘 丽, 邢美园. 单篇论著影响力评价指标比较分析——学术迹与 Altmetrics 评分、F1000 评分、Comment 的比较[J]. 情报理论与实践, 2017, 3(40):114-118.
- [27] 金贞燕, 侯景丽, 孙华丽. Altmetrics 数据整合分析工具的现状特点及相关问题研究[J]. 情报理论与实践, 2019, 42(4): 89-95.
- [28] Al B J P I. Peer review, bibliometrics and altmetrics – do we need them all: lydia s. middleton[C]. Association for Information Science and Technology. 81st Annual Meeting of the Association for Information Science & Technology, Maryland, 2018:653-656.
- [29] Mongeon P. Using social and topical distance to analyze information sharing on social media[C]. Association for Information Science and Technology. 81st Annual Meeting of the Association for Information Science & Technology, Maryland, 2018: 397-403.
- [30] DePaula N, Million A J, Fietkiewicz K J, et al. Challenges for social media: misinformation, free speech, civic engagement, and data regulations[C]. Association for Information Science and Technology. 81st Annual Meeting of the Association for Information Science & Technology, Maryland, 2018:665-668.
- [31] Lemke S, Mazarakis A, Mehrazar M, et al. Are There different types of online research impact?[C]. Association for Information Science and Technology. 81st Annual Meeting of the Association for Information Science & Technology, Maryland, 2018: 282-289.
- [32] Gorichanaz T. Art and everyday information behavior: sources of understanding[C]. Association for Information Science and Technology. 81st Annual Meeting of the Association for Information Science & Technology, Maryland, 2018:143-150.
- [33] Zhou Q, Sin S J, Lee C S. Beyond mandatory use: probing the affordances of social media for formal learning in the voluntary context[C]. Association for Information Science and Technology. 81st Annual Meeting of the Association for Information Science & Technology, Maryland, 2018:608-617.
- [34] Bornmann L, Haunschild R, Adams J. Do altmetrics assess societal impact in a comparable way to case studies? An empirical test of the convergent validity of altmetrics based on data from the UK Research Excellence Framework (REF)[J]. Journal of Informetrics, 2019, 13(1):325-340.
- [35] Ortega J L. The life cycle of altmetric impact: a longitudinal study of six metrics from PlumX[J]. Journal of Informetrics, 2018, 12(3):579-589.
- [36] 余厚强, 曹雪婷. 替代计量数据质量评估体系构建研究[J]. 图书情报知识, 2019(2):19-27.
- [37] 田文灿, 胡志刚, 王贤文. 科学计量学视角下的 Altmetrics 发展历程分析[J]. 图书情报知识, 2019(2):5-11.
- [作者简介] 牌艳欣, 女, 1996 年生, 山东理工大学科技信息研究所硕士研究生。
- 李长玲, 女, 1969 年生, 山东理工大学科技信息研究所教授, 硕士生导师。
- 刘运梅, 女, 1995 年生, 山东理工大学科技信息研究所硕士研究生。
- 收稿日期: 2019-03-25

## 欢迎订阅

### 《复印报刊资料·出版业》

精选、转载出版理论研究的优秀学术成果;关注国家出版工作的方针政策;反映出版业界的改革与发展;研讨出版实践中的前沿问题,理论性与应用性并重,是出版界同仁交流、学习的高端参考读物。

栏目主要有:发展战略、学术平台、业界观察、数字出版、出版法苑、书业经纬、期刊纵横、编校实务、出版史论、环球出版、探索与争鸣等。

月刊,大 16 开,80 页,国内外公开发行。每期定价 19.00 元,全年定价 228 元。

编辑部地址:北京市海淀区中关村大街甲 59 号文化大厦 1301 室

编辑部电话:010-82504550-2165

营销热线:010-82503438、82503439、82503412、82503029

网址:www.zlzx.org