

# 学科交叉视角下的网络舆情研究主题比较分析 ——以国内图书情报学和新闻传播学为例

王连喜, 曹树金

(中山大学资讯管理学院, 广州 510006)

**摘 要** 从学科交叉视角探索网络舆情在图书情报学和新闻传播学中的研究共性和差异性。以 CNKI 数据库中两门学科核心期刊的网络舆情研究文献为样本, 在关键词规范化的基础上, 一方面从微观角度结合词频和 TF-IDF 方法提取重要关键词, 并对两门学科的关键词进行交集分析、共词聚类分析和战略坐标分析; 另一方面从全局角度采用 LDA 模型对样本进行主题建模。然后, 分别对基于共词聚类和基于 LDA 的热点主题划分结果进行语义关系规范和主题标识。研究发现, 舆情传播与监测、高校舆情分析、危机预警与管理、政府舆情应对等是两门学科共同关注的主题, 网络谣言分析、网络问政则分别是图书情报学和新闻传播学的特色研究主题。研究还发现, 有部分潜在交叉主题和衍生主题在学科交叉与融合的推动下孕育而生。

**关键词** 网络舆情; 热点主题; 学科交叉; LDA; 共词聚类

## Comparative Analysis on the Hot Topics of Network Public Opinion from the Research of Cross-Disciplinary Perspective: An Example of Library & Information Science and Journalism & Communication

Wang Lianxi and Cao Shujin

(School of Information Management, Sun Yat-Sen University, Guangzhou 510006, China)

**Abstract:** To explore the common and difference about the network public opinion between Information Science and Journalism & Communication from the research of cross-disciplinary perspective. Using CNKI as the data resource, this paper selects the papers with title, keywords and abstract about the network public opinion in the core journals of the two disciplines. On the basis of normalized keywords, this work combines the words frequency and TF-IDF to extract primary keywords and analysis the intersection between them. In additional, it also uses co-word clustering, strategic diagrams and LDA to analyze the hot topics for the two disciplines, and normalizes the semantic and identifies the topics for clusters. The results show that the hot topics, such as network public opinion propagation and monitoring, college and university network public opinion analysis, crisis early warning and management, government response, are the common focus on the two disciplines, and the topics of internet rumors and network politic management are the unique research topics in Information Science and Journalism & Communication, respectively. Furthermore, some potential crossing topics and derivative topics are produced by the integration and intercrossing between disciplines.

**Key words:** network public opinion; hot topics; discipline-crossing; LDA; co-word clustering

收稿日期: 2016-05-24; 修回日期: 2016-09-17

基金项目: 国家社会科学基金重大项目(12&ZD222), 国家自然科学基金项目(61572145), 教育部人文社会科学研究青年项目(14YJC870021, 13YJCZH258), 广东省科技计划项目(2014A040401083, 2015A030401093)。

作者简介: 王连喜, 男, 1985 年生, 中山大学资讯管理学院博士研究生, 主要研究方向为自然语言处理与文本挖掘; 曹树金, 男, 1962 年生, 博士, 中山大学资讯管理学院教授, 博士生导师。

舆情是一个具有中国特色的抽象化概念,它与舆论是一对从属性的概念,相互之间既有共性又有差异<sup>[1]</sup>。从发展起源上看,舆情与英文词汇 Public Opinion 所表达的范畴既有交叉又有重叠,所以在中文表述方面常与“舆论”、“民意”和“公众意见”等词汇互通互用。曾润喜认为舆论是多数人形成的一致性意见,只有当持有某种意见的人数达到一定的量才能成为舆论,而舆情是人们的认知、态度、情感和行为倾向的原初表露,可以是一种零散的、非体系化的东西,也不需要得到多数人认同,是多种不同意见的简单集合,所以当舆情产生聚集时就可以向舆论转化<sup>[2]</sup>。高承实等<sup>[3]</sup>认为网络舆情的起点始于网络舆情事件发生,终点止于网络舆论形成,或现实世界中占主导地位的针对性行动出现,或舆情消亡。由此可知,舆情是舆论的初始表现,是舆论形成的前期准备过程。

自2004年中国共产党十六届四中全会提出建立和完善舆情信息汇集和分析机制的战略意见以来,网络舆情研究进入了快速的发展阶段,并得到了信息科学、社会学、政治学、心理学、图书情报学、新闻传播学和管理学等跨学科、跨领域学者的关注<sup>[4-6]</sup>。目前,网络舆情研究主要涉及信息科技领域、新闻传播学领域和交叉学科领域。信息科技领域包括计算机、人工智能、图书情报等学科,该领域研究人员主要从技术视角针对网络舆情信息的采集、处理、分类、监测、预警等技术的研发和管理进行深入研究,但是由于过分重视技术方法而缺乏多元视角或全局性的分析;新闻传播学领域学者大多从传播学或舆论宣传视角着重关注网络舆情的内容管理和信息引导,容易偏离中国社会现实;而交叉学科领域学者则在综合技术和内容研究的基础上,引导网络舆情研究向其他领域渗透,但是存在着理论支撑和研究滞后于中国社会舆情发展现状的问题,同时也存在着学科边界模糊的问题<sup>[7,8]</sup>。近年来,一些研究表明不同学科在宏观层面上具有很强的关联关系和交叉关系<sup>[9-12]</sup>。例如,闵超和孙建军从定量角度以系统聚类分析法和战略坐标分析法对图书情报学和新闻传播学的关键词交集进行研究,发现网络舆情已在近几年成为了两门学科共同关注的焦点<sup>[12]</sup>。

本研究试图从图书情报学科和新闻传播学科的核心期刊题录数据入手,以文献计量法、文本挖掘和内容分析法为手段,深入探索网络舆情研究在两门学科领域的共性和差异性,明晰其热点主题的分

布特点和结构,揭示国内网络舆情研究在两门学科领域的研究热点和特色化发展趋势。

## 1 研究思路与方法

不同学科之间研究内容的交叉与联系通常可以用一段时期内学术研究成果的关键词集合来揭示。如果两门学科在某个研究领域上存在越多的具有相似语义关系的关键词分布,说明两者之间具有较大的研究共性,越容易产生新的学科增长点和创新力;如果两个学科在相同研究主题下存在较多的交叉关键词,说明两者在研究内容或研究方法方面具有更多的相似之处,越容易产生新的创新潜力<sup>[13]</sup>。基于该理论假设,本研究对网络舆情研究在图书情报学和新闻传播学的关键词集合进行交集分析、词频分析、共词聚类分析、战略坐标分析以及主题挖掘分析,最后透视出两个学科在该领域上的学科交叉点和新的知识增长点。

### 1.1 共词聚类分析

共词聚类分析是一种常见的文献内容分析方法。它首先通过统计高频关键词在研究对象中两两共同出现的频率,然后对高频词进行特征向量表示并形成共词矩阵,并以相似性度量或距离度量方法对关键词进行聚类划分,使词对共现频率较高的主题词聚集成簇,最后依据簇中关键词的语义关系归纳出该簇所表征的内容。

共词聚类与其他文本聚类算法一样,在特征表示之前需要筛选出能够表征研究对象的关键知识点。但是许多研究表明,共词聚类的效果与所选关键词的质量有非常大的关系,容易受到局部性核心关键词的影响。因此,部分研究人员通过对传统度量方法进行改进以突出具有实际意义但词频相对较低的关键词。例如,胡昌平等<sup>[14]</sup>针对高频词矩阵会丢失大量重要共现关系的现象,提出利用关键词增补方法来优化高频词矩阵。王连喜等<sup>[15]</sup>通过结合TF-IDF方法和高频低频词分界定律来选择显性和隐性的重要关键词。

通常,在共词聚类完成之后可采用黏合力来度量簇中各个特征对聚类结果的贡献程度,并按照黏合力大小及语义关系对每个簇的结果进行语义关系规范和主题命名。簇的黏合力可以用该簇内各个特征词与其他特征词共现频次的均值来度量。黏合力

大的特征词为簇的中心词，对确定主题标签起着非常重要的作用。因此，簇的主题标签可用簇中黏合力大的特征词来标识。

### 1.2 战略坐标分析

战略坐标是共词聚类结果的一种可视化形式，旨在可视化图形中体现簇内关系和簇间联系。战略坐标图的本质就是一张二维坐标图，以向心度和密度为坐标轴形成 4 个象限，其基本结构是横轴表示向心度，纵轴表示密度，原点为两个轴的中位数或平均值(如图 1 所示)。X 轴的向心度可以用来测量不同簇之间的相互关系，Y 轴的密度可以用来测量相同簇内部联系的强弱。对于某一学科的研究主题而言，如果向心度越大，说明该主题与其他主题关联非常紧密，显性表现为两个簇中关键词共现的频率较高；如果密度越大，表明该主题研究渐趋成熟，显性表现为相同簇中不同关键词的共现频率较高。



图 1 战略坐标图

从核心边缘程度的角度看，第 I 象限中的簇属于该研究领域核心内容，表现为研究热点；第 II 象限中的簇可能是该研究领域较为成熟的研究内容，但或许尚未成为该领域的研究核心；第 III 象限中的簇属于发展不成熟，结构不稳定的研究内容；第 IV 象限的中簇为该研究领域核心研究内容，但该主题可能发展还不够成熟。

### 1.3 LDA 主题挖掘模型分析

LDA 主题模型是一种常用的文本主题挖掘方法，它是一个具有三级层次结构的贝叶斯模型，包括“特征-主题-文档”三个层次<sup>[16]</sup>。LDA 模型将隐含主题看做成词语特征的概率分布，同时将每个文档看做是潜在主题的概率分布<sup>[17]</sup>。因此，利用 LDA 模型能够挖掘隐藏在文本内不同主题与特征词之间的关系，从而得到文本的主题分布，同时也能得到

文本  $doc_i$  属于每个主题  $z_j$  的一个概率值  $p\left(\frac{doc_i}{z_j}\right)$ ,

即特征  $f_j$  的权重为  $w_{ij} = p\left(\frac{doc_i}{z_j}\right)$ 。

LDA 主题模型是一种从全局角度对文档、主题和关键词进行概率分布表示的聚类方法，所以它能够在不需要机器真正理解自然语言的情况下随机生成具有观测性的数据，并能够提取出可以被理解的、相对稳定的隐含语义结构，从而简化文本的主题表达。

## 2 数据获取与处理

### 2.1 数据来源与获取

选取 CNKI 收录的图书情报学和新闻传播学领域核心期刊所发表的网络舆情相关论文为样本。由于网络舆情不是档案学和出版学的研究范畴，因此将两者分别从图书情报学和新闻传播学的期刊范围中予以剔除。另外，由于《情报学报》不在 CNKI 的收录范围内，所以需要额外增补。

基于检全率和检准率考虑，将检索条件设为主题、文献来源限定为两门学科的核心期刊、时间限定为 2016 年 9 月 1 日之前所发表的文献，以“网络舆情”为检索词进行检索。通过对检索后的文献进行去重、清洗、增补等步骤后共得到有效结果为 972 篇(包括增补的《情报学报》题录数据)，其中图书情报学 480 篇，新闻传播学 492 篇。从图 2 所示的年发文量的统计结果看，从 2009 开始两门学科关于网络舆情的研究出现非常明显的增长趋势，但是从 2014 年开始进入平缓期。图 3 显示，图书情报学中关于网络舆情的文章主要刊载于情报学期刊上(约占该学科发文总量的 90%)，包括《情报杂志》、《情报

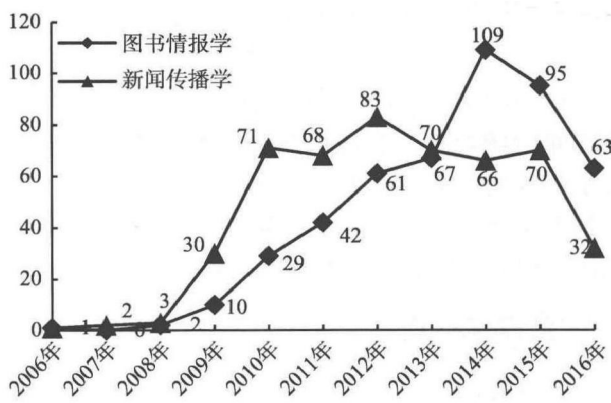


图 2 中国网络舆情研究学术关注度

学报》、《图书情报工作》、《情报科学》、《情报理论与实践》、《图书情报知识》、《情报资料工作》。图 4 显示,新闻传播学中关于网络舆情的文章主要刊载于《青年记者》、《新闻战线》《新闻记者》、《新闻爱好者》、《中国记者》、《新闻与写作》、《当代传播》等期刊。

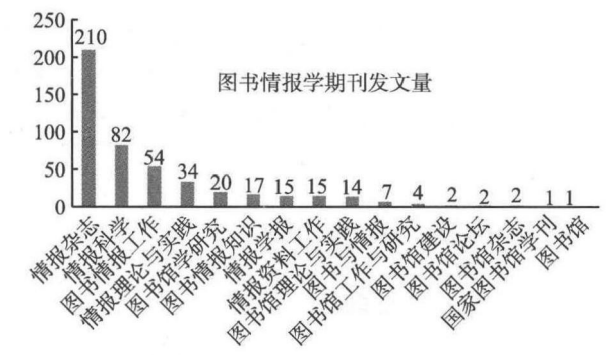


图 3 图书情报学期刊发文量

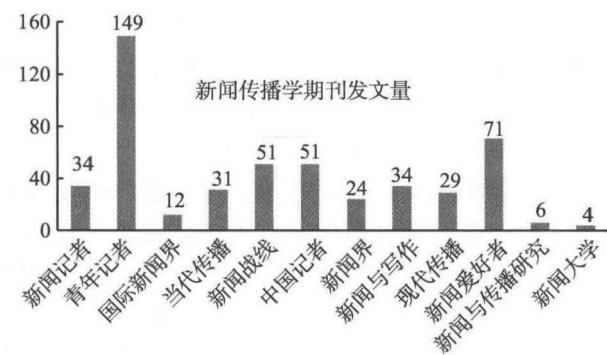


图 4 新闻传播学期刊发文量

此外,为了揭示不同研究机构在该研究领域上的影响力,本研究分别对图书情报学和新闻传播学的机构发文量进行统计,其排名前十的机构如表 1 所示。结果显示,具有信息科学、公共管理、经济管理 etc 学科性质的研究机构比较喜欢向图书情报学领域期刊投稿,而新闻传播学领域的期刊则主要刊载具有传播学或舆论宣传视角研究机构的网络舆情论文。

统计结果还表明,重庆大学新闻学院、南京大学新闻传播学院、北京大学信息管理系等机构都有跨学科投稿的尝试,但是它们的发文量在两门学科中都不是很大。可见,部分研究机构在网络舆情领域已出现了多学科交叉的研究趋势。

2.2 数据预处理

将下载的检索文献题录数据进行预处理,包括数据合并、数据规范化转换、数据清理等步骤。在

数据规范化转换过程中,一方面对关键词进行同义词、近义词合并,另一方面对具有上下位关系的关键词进行归并处理(如表 2 所示)。

表 1 两学科发文量排名前十的机构

图书情报学		新闻传播学	
第一机构	数量	第一机构	数量
华中科技大学公共管理学院	30	中国人民大学新闻学院	18
吉林大学管理学院	16	湖南大学新闻传播与影视艺术学院	8
南京大学信息管理学院	16	暨南大学新闻与传播学院	8
福州大学公共管理学院	15	人民网舆情监测室	7
中国人民武装警察部队学院	15	复旦大学新闻学院	5
福州大学经济与管理学院	12	南京师范大学新闻与传播学院	5
北京邮电大学经济管理	11	中国传媒大学电视与新闻学院	5
武汉大学信息资源研究中心	9	大众网	4
新疆财经大学统计与信息学院	7	河南工业大学新闻与传播学院	4
南京邮电大学管理学院	6	华中科技大学新闻与信息传播学院	4

同时,为了能够清晰地识别出标题和摘要中的关键词,本研究需要对题录数据进行分词。首先将作者标识的关键词导入 ICTCLAS 分词系统的自定义词典中,并将无意义的通用词或泛化的词语添加到分词系统的停用词列表中,然后将规范化后的数据整理成文本训练集,并对其进行分词,最后统计分词后关键词的词频及其首次出现的时间。

2.3 共词矩阵构建

考虑到高频词是表征热点研究主题的显性范式,但是由于在短时间内能够体现最新研究主题的词频较低,往往容易被忽略,所以本研究结合关键词词频及 TF-IDF 值提取重要关键词。

TF-IDF 是一种词权重计算方法,其主要思想是:如果一个关键词在一篇文档中出现的频率高,并且该词在其他文档中很少出现,则认为是重要关键词。TF-IDF 的常用表示形式如下:

$$TF-IDF = \frac{\frac{n_{i,j}}{N_i} \times \log \frac{N}{m_{i,j}}}{\sqrt{\sum_{j=1}^m \left( \frac{n_{i,j}}{N_i} \times \log \frac{N}{m_{i,j}} \right)}}$$

其中，表示  $n_{i,j}$  词  $t_{i,j}$  在文档  $i$  中出现的次数， $N_i$  表示文档  $i$  中所有词语出现的次数， $N$  表示文档信的总个数， $m_{i,j}$  表示包含该词语的文档个数。

表2 关键词规范化（部分）

规范化后	规范化前
舆论引导	网络舆论引导、舆论引导工作、舆论引导力、舆论引导作用、舆论引导功能、政府舆论引导
公众情绪	情绪化、民众情绪、社会情绪、表达情绪
议程设置	议题设置、议程设置理论、媒介议程设置、议程设置功能
突发事件	非常规突发事件、突发公共事件、公共危机事件、高校突发事件、网络突发事件、校园突发事件
网络谣言	微博谣言、谣言演化、谣言净化者、谣言事件、谣言传播、谣言治理
舆情监测	信息监测、网络舆情监控、舆情监测系统、舆情监管、网络舆情监测
聚类	聚类分析、文本聚类、K-Means 算法、话题聚类
情感分析	情感倾向分析、文本倾向性、情感倾向
应急决策	应急管理、应急情报、舆情应急
共词分析	共词分析法、共词网络、共现特征词、共现分析
文献计量	信息计量学、文献计量法、信息计量、文献计量学
热点发现	热点跟踪、热点话题追踪、热点话题识别、热点主题
可视化	可视化分析、可视化建模、信息可视化
观点挖掘	观点抽取、观点演化、互联网舆情观点、舆情观点
指标体系	监测指标、评价指标、评价指标体系、舆情指标、预警指标、预警指标体系、指标设计
意见领袖	网络意见领袖、舆论领袖、新意见领袖、微博意见领袖
社会网络	社会网络分析法、社交网络、舆情网络、用户社会网络、社会网络挖掘、SNS
危机处理	危机处置、危机干预、危机预警机制、危机应对、危机报道、危机公关、危机传播管理、危机反应策略

通过 TF-IDF 计算结果确定两门学科的重要关键词并过滤掉一些意义不大的关键词，然后结合 Donohue 提出的高频低频分界定律，最终确定图书情报学的重要关键词为 60 个，新闻传播学的重要关键词为 63 个。接下来，针对两门学科的重要关键词分别构建 60×60 和 63×63 的共词矩阵。共词矩阵非对角线上的元素为两两关键词在同一篇文献中出现的次数，对角线上元素为每一行或第一列中的最大值加 1。

然而，针对“同量不同质”的问题，本研究采用改进的 Ochiai 系数法将共词矩阵转换成相异矩阵（表 3 和表 4 所示），以便体现出共词矩阵中同量数据的差异性。改进的 Ochiai 系数计算方法如下：

$$Ochiai = 1 - \left( \frac{c_{ij}}{\sqrt{C_i \cdot C_j}} \right) \cdot \log(\log(C_{ij})),$$

其中， $C_{ij}$  表示关键词  $i$  和  $j$  共现的次数， $C_i$  是关键词  $i$  出现的频次， $C_j$  是关键词  $j$  出现的频次。

表3 图书情报学相异矩阵（部分）

	网络舆情	突发事件	微博	舆情	指标	监测	社会网络
网络舆情	0.00	0.762	0.866	0.677	0.837	0.764	0.891
突发事件	0.762	0.00	0.870	0.890	0.928	0.938	0.984
微博	0.866	0.870	0.00	0.900	0.811	0.949	0.733
舆情	0.677	0.890	0.900	0.00	0.949	0.905	0.993
指标	0.837	0.928	0.811	0.949	0.00	0.942	0.876
监测	0.764	0.938	0.949	0.905	0.942	0.00	0.952
社会网络	0.891	0.984	0.733	0.993	0.876	0.952	0.00

表4 新闻传播学相异矩阵（部分）

	网络舆情	舆情	微博	新媒体	突发事件	网络舆论	政府
网络舆情	0.00	0.502	0.777	0.705	0.810	0.725	0.698
舆情	0.502	0.00	0.886	0.820	0.829	0.846	0.820
微博	0.777	0.886	0.00	0.871	0.848	0.897	0.830
新媒体	0.705	0.820	0.871	0.00	0.991	0.935	0.870
突发事件	0.810	0.829	0.848	0.991	0.00	0.889	0.844
网络舆论	0.725	0.846	0.897	0.935	0.889	0.00	0.874
政府	0.698	0.820	0.830	0.870	0.844	0.874	0.00

## 3 研究结果分析

### 3.1 交叉关键词分析结果

为了尝试从微观角度探索图书情报学和新闻传播学在网络舆情研究方面存在的共性和差异性，本研究在关键词词频计算的基础上，统计两门学科关键词的首次出现时间及两门学科关键词集合的交集。表 5 给出了网络舆情研究在图书情报学（A）和新闻传播学（B）中出现的特色关键词频度及其首次出现时间，以及两学科共同出现的关键词的频度及其首次出现时间。由表 5 所示结果可知，两门学科在网络舆情研究方面都聚焦于微博、论坛等网络媒体的突发事件、热点事件、群体性事件的发现及其内容分析、意见领袖识别和网络谣言检测等内容，但是在关注时间上新闻传播学要比图书情报学早。

综观两门学科的特色关键词可以发现，图书情报学从 2010 年开始主要偏向于网络舆情的信息处理技术和方法研究，而新闻传播学在前期重点关注网络舆论的传播、网络问政、公众情绪引导等内容，

表5 两学科关键词词频及其首次出现时间(部分)

特色关键词(A)	特色关键词(B)	交叉关键词(A:B)
指标体系(124, 2008)	舆论监督(69, 2009)	网络舆情(1049, 2006; 781, 2006)
社会网络(127, 2010)	舆论场(63, 2009)	突发事件(246, 2009; 196, 2009)
预警(103, 2009)	网络问政(41, 2010)	微博(243, 2010; 274, 2011)
文献计量(26, 2011)	网络民意(28, 2009)	意见领袖(68, 2010; 120, 2009)
情报学(45, 2011)	媒介素养(27, 2010)	大数据(65, 2013; 43, 2013)
情感分析(41, 2012)	社会政治态度(23, 2007)	网络谣言(63, 2012; 41, 2010)
聚类(55, 2013)	公众情绪(22, 2007)	群体极化(38, 2010; 18, 2009)
系统动力学(38, 2010)	躲猫猫(16, 2009)	内容分析(32, 2012; 4, 2009)
观点挖掘(37, 2009)	传播学研究(15, 2008)	知识图谱(17, 2011; 6, 2015)
竞争情报(21, 2012)	新华网(13, 2009)	热点事件(17, 2011; 37, 2009)
电子政务(21, 2012)	把关人(12, 2006)	舆论引导(15, 2014; 128, 2009)
神经网络(21, 2011)	舆论生态(11, 2012)	群体性事件(14, 2010; 61, 2010)
LDA(19, 2012)	蝴蝶效应(10, 2010)	沉默的螺旋(10, 2012; 14, 2013)
热点发现(6, 2012)	媒介融合(9, 2009)	议程设置(9, 2014; 57, 2006)

但从2010年之后则开始与图书情报学的研究知识点出现交叉性。

### 3.2 共词聚类结果

为了清晰地呈现出网络舆情研究的热点主题或子知识点,在导入SPSS之前先将相异矩阵的首行和首列删除,以便消除研究对象对该领域关键词的聚类效果。在进行系统聚类时,以相异矩阵为输入,以Ward法、欧氏平方距离和Z得分标准化为参数设置,得到两门学科的共词聚类结果(如图5和图6所示)。聚类的实质是“物以类聚”,它可将具有相同类型或相近语义关系的关键词聚集成簇。因此,从关键词的共现关系以及聚类效果的角度看,图5和图6所示的树状图结果可分别划分为10个簇和8个簇。

另外,为了清晰地展示关键词的作用和目的,本研究以黏合力为指标确定各簇的核心关键词和簇标签,同时依据文献[14]提出的语义标注方法将关键词归纳为能够揭示所在簇的研究主题、或标明研究的所属领域、或限定研究成立的范围、或概括研究涉及的知识点、或描述研究依据的理论或方法等五种语义类型。表6和表7是分别根据图5和图6所示的聚类结果所归纳出的图书情报学和新闻传播学关键词聚类结果和每个关键词在所属簇中的黏合力大小及其语义类型。

从聚类划分的结果中可以看出:(1)不同主题簇的关键词数量和核心关键词的黏合力大小有所差异;(2)关键词数量多的主题簇,平均黏合力小;(3)网络舆情预警与监测、突发事件舆论引导、危

机处理等三个研究主题在两门学科中都被单独聚集成簇,说明这两个主题已成为两门学科研究人员在交叉领域的关注重点;(4)从语义类型的标注上看,图书情报学主要偏向于网络舆情信息处理及学习的理论和方法研究,而新闻传播学则多关注某个特定领域或范围内的舆情监测或舆论传播研究。

当然,在两门学科的部分主题簇中存在着相同或相近关键词,但是这些关键词在所属簇中的黏合力大小有差异,而且它们与所属簇的其他关键词的语义关系也不一样,所以导致其所属簇的主题标签不相同。上述分析表明,关键词的表征范围和关键词的选择会对共词聚类的结果及其命名和解读产生一定的影响<sup>[18]</sup>。

### 3.3 战略坐标分析结果

在得到共词聚类结果之后,根据相关矩阵计算两门学科关键词聚类划分结果的向心度和密度,然后分别以两门学科所有簇的向心度和密度的均值作为战略坐标图的原点,并以各个簇的向心度和密度与原点的差值作为该簇在战略坐标图上的坐标,最终形成的战略坐标图如图7和图8所示。

从主题分布上看,图书情报学横轴上方的主题数量分布比新闻传播学多,但两门学科在纵轴右侧的主题数量分布相当。在两门学科的战略坐标图中,纵轴左侧和横轴下方区域集中了大多数主题,说明这些主题的关键词共现频度相对较低,发展不够成熟,其原因在于这几个主题是几年新出现的,刚刚

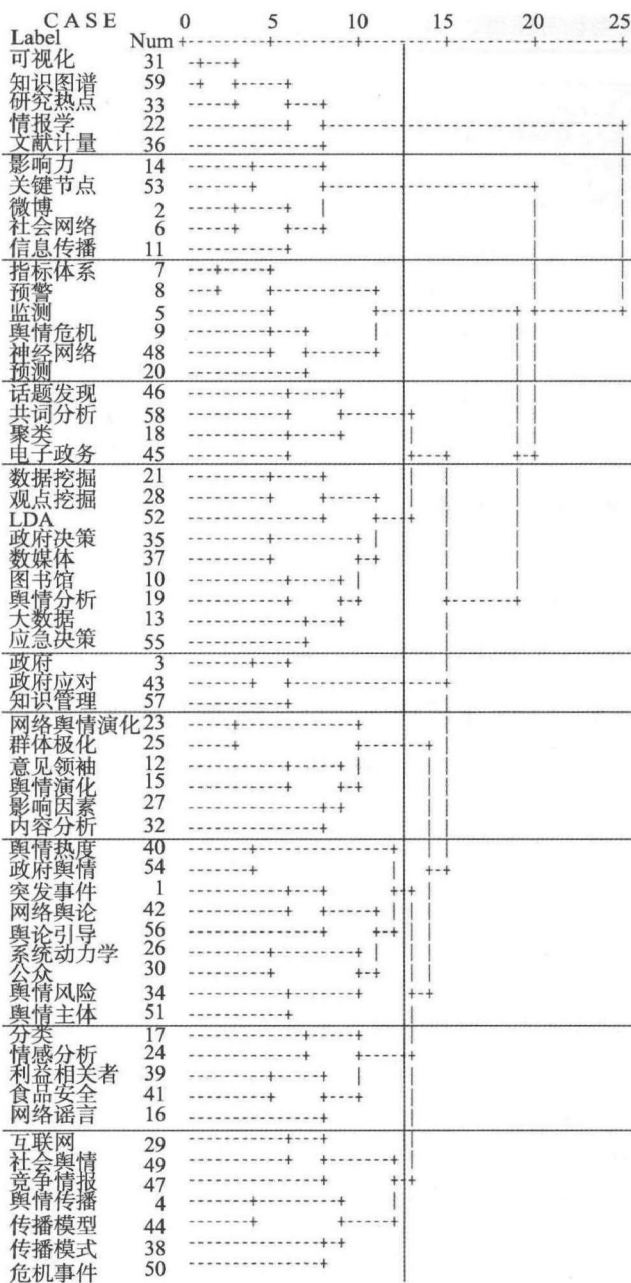


图5 图书情报学共词聚类结果

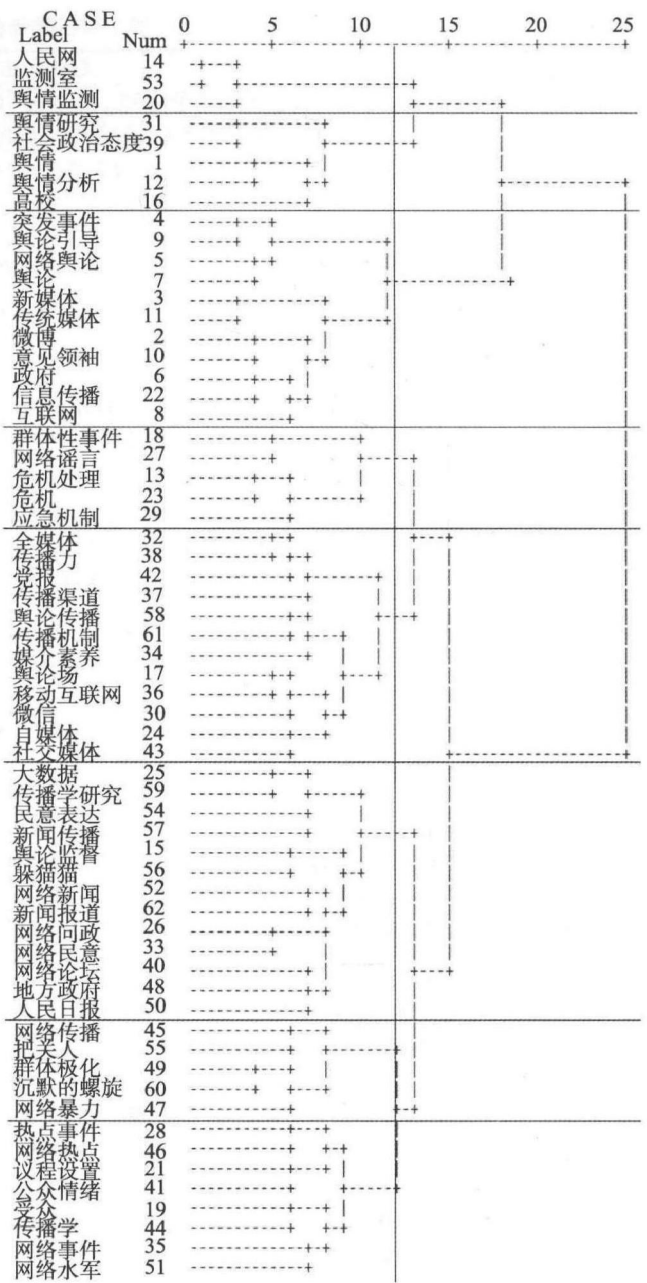


图6 新闻传播学共词聚类结果

表6 图书情报学聚类划分结果

簇标签	关键词/黏合力	关键词语义类型
研究热点分析	研究热点/4.8、可视化/6、知识图谱/5.2、文献计量/3、情报学/4.6	研究主题、所属领域、限定范围
节点影响力分析	微博/19.8、社会网络/15、信息传播/10.8、影响力/11.6、关键节点/4.6	研究主题、理论方法、子知识点
网络舆情预警	监测/9.5、指标体系/12.67、预警/15.67、舆情危机/6.83、预测/3.67、神经网络/4.33	研究主题、理论方法、子知识点
舆情话题发现	舆情演化/0.5、电子政务/0.3、话题发现/1.25、共词分析/0.75	研究主题、理论方法、子知识点
大数据舆情分析	图书馆/1、大数据/1.56、舆情分析/2.78、数据挖掘/2.33、观点挖掘/1.56、政府决策/1.67、新媒体/1.22、LDA/0.778、应急决策/0.44	所属领域、限定范围、理论方法、子知识点
政府舆情应对	政府/8、政府应对/6.33、知识管理/4.33	所属领域、研究主题、子知识点
网络舆情演化	意见领袖/3.17、舆情演化/4、网络舆情演化/5、群体极化/3.3、影响因素/2.5、内容分析/1	研究主题、子知识点、理论方法
突发事件舆论引导	突发事件/5、系统动力学/2.44、公众/1.67、舆情风险/1、舆情热度/1.89、网络舆论/2、舆情主体/1.56、政府舆情/0.89、舆论引导/0.67	研究主题、限定范围、子知识点、理论方法
网络谣言分析	网络谣言/2、分类/2.8、情感分析/1、利益相关者/1.8、食品安全/2	限定范围、理论方法、子知识点
危机事件舆情传播	舆情传播/4.29、互联网/1.429、传播模式/1.143、传播模型/2.15、竞争情报/0.58、社会舆情/1.58、危机事件/0.58	研究主题、子知识点、限定范围



表7 新闻传播学聚类划分结果

簇标签	关键词/黏合力	关键词语义类型
舆情监测	人民网/11.33、舆情监测/9.33、监测室/8.67	研究主题、子知识点
高校舆情分析	舆情/19.4、舆情分析/8.8、高校/6.8、舆情研究/6.4、社会政治态度/6.2	所属领域、限定范围、子知识点
突发事件舆论引导	微博/22.82、新媒体/19.82、突发事件/16.46、网络舆论/20.82、政府/19、舆论/17.64、互联网/15.27、舆论引导/18.64、意见领袖/15.36、传统媒体/12.09、信息传播/9.91	限定范围、研究主题、子知识点
群体性事件危机处理	危机处理/5、群体性事件/4.2、危机/5、网络谣言/2、应急机制/3.4	研究主题、限定范围、子知识点
舆论传播	舆论场/3.42、自媒体/2.08、微信/2.42、全媒体/1.92、媒介素养/1.08、移动互联网/2.08、传播渠道/1.75、传播力/1.42、党报/0.92、社交媒体/0.92、舆论传播/1、传播机制/1	研究主题、限定范围、子知识点
网络问政	舆论监督/0.69、大数据/0.38、网络问政/1.31、网络民意/1.62、网络论坛/1.15、地方政府/0.46、人民日报/0.54、网络新闻/0.46、民意表达/0.62、躲猫猫/0.46、新闻传播/0.46、传播学研究/0.62、新闻报道/0.46	研究主题、所属领域、限定范围、子知识点、理论方法
网络传播控制	网络传播/1.4、网络暴力/1.2、群体极化/1.8、把关人/1、沉默的螺旋/2.2	研究主题、限定范围、子知识点
议程设置	受众/2.63、议程设置/2.63、热点事件/2.5、网络事件/1.25、公众情绪/1.25、传播学/1.63、网络热点/1.38、网络水军/1.25	研究主题、限定范围、所属领域、子知识点

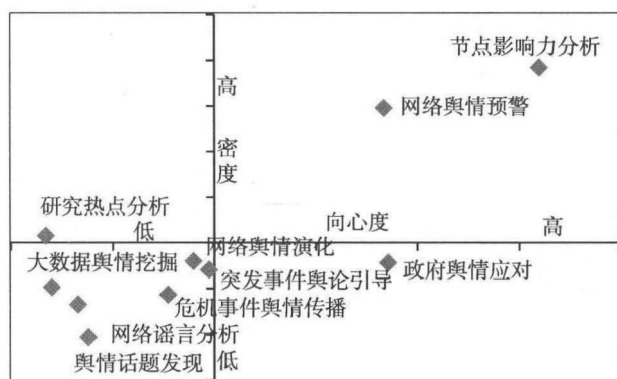


图7 图书情报学研究主题的战略坐标图

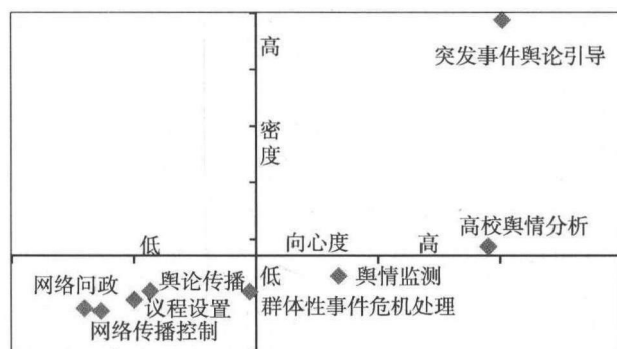


图8 新闻传播学研究主题的战略坐标图

被领域内的专家关注时间不长。这个结果与关键词首次出现时间的统计结果相吻合。

从主题的核心程度上看，两门学科对于网络舆情监测与预警、突发事件舆论引导两个主题的关注时间都比较长，发展得非常成熟且稳定，但两者的侧重点有所不同，图书情报学关注突发事件的舆情评估与引导，而新闻传播则聚焦于群体性事件舆情

的传播与引导。

### 3.4 LDA 主题挖掘与语义揭示结果

为了弥补局部性关键词的选择对共词聚类造成部分共现关系丢失和新兴关键词缺失的问题，本研究从全局视角出发，首先对题录数据进行分词，然后对分词后得到的关键词采用LDA模型进行主题建模。基于LDA的主题模型利用词对共现频率进行词组聚类，旨在将词向量空间转化为主题空间，通过主题降维得到的主题概率分布往往具有语义特征，因此可以更有效地将主题相近的词或词组聚成一类。在实验中，设吉布斯抽样的主题数目 $T=10$ ，将超参数设置为 $\alpha=5$ 、 $\beta=0.1$ 、 $\text{Iteration}=100$ 。在分别对两门学科的数据进行LDA建模之后，得到如表8和表9所示的10个潜在主题的概率分布。表8和表9中每主题下的前10个重要关键词按权重值大小进行降序排列，而主题标签则是依据各个主题中重要主题词的语义关系所归纳出来的。

由表8和表9所示结果可知，基于LDA的主题挖掘结果与共词聚类结果大致相同，但是同一学科相同主题下的重要关键词分布会有所差异。相比之下，基于LDA的主题挖掘结果中关键词的数量更多，其语义联系更为紧密，内容描述也更加具体和明确。然而，由于基于共词聚类的主题划分方法和基于LDA的主题挖掘方法是从不同视角进行关键词选择的，且两者的研究方法也是基于不同的关系而进行的主题划分，所以两者的主题挖掘结果是可以形成互补的。因此，综合两种主题划分结果则可以分别得到两门学科的网络舆情研究热点。



表8 图书情报学科的 LDA 主题模型挖掘结果

主题标签	前 10 个重要主题词及权重
政府舆情应对	政府/0.521、舆情/0.063、系统动力学/0.050、舆情风险/0.044、知识管理/0.039、信息服务/0.028、舆情热度/0.027、网络推手/0.017、电子政务/0.008、谣言传播/0.008
网络舆情监测	网络舆情/0.453、监测/0.227、意见领袖/0.100、舆情分析/0.082、移动环境/0.017、微分方程/0.014、知识基础/0.014、社会舆情/0.008、公共危机/0.008、群体性事件/0.008
突发事件舆情传播	突发事件/0.406、指标体系/0.230、信息传播/0.129、层次分析法/0.042、网络媒体/0.025、网络论坛/0.017、安全评估/0.017、议程设置/0.015
网络舆情预警	网络舆情/0.535、预警/0.234、利益相关者/0.048、舆情/0.032、元胞自动机/0.025、文献计量/0.024、话题演化/0.018、社会舆情/0.014、引导机制/0.010、社交媒体/0.006
大数据舆情传播	网络舆情/0.306、舆情传播/0.251、大数据/0.098、舆情演化/0.089、舆情研究/0.033、政府公信力/0.020、应对策略/0.017、演化博弈/0.017、传播模型/0.017、政府决策/0.012
基于情报学的舆情分析	微博/0.542、情报学/0.075、研究热点/0.054、内容分析/0.051、可视化/0.046、LDA/0.031、网络舆论/0.026、舆论引导/0.021、知识图谱/0.018、前沿主题/0.015
网络舆情危机处理	社会网络/0.219、网络舆情/0.214、舆情危机/0.148、群体极化/0.055、舆情事件/0.055、影响因素/0.051、食品安全/0.043、政府决策/0.031、关键节点/0.025、社会影响/0.013
网络舆情演化分析	舆情演化/0.761、图书馆/0.103、互联网/0.020、网络结构/0.014、网络评论/0.014、数值仿真/0.010、实证研究/0.010、情报学/0.006、案例分析/0.004、政府/0.003
高校舆情分析	舆情/0.137、高校/0.055、观点挖掘/0.049、热点事件/0.030、情感分析/0.030、实证研究/0.021、群体性事件/0.019、危机预警/0.011、评价指标/0.008、意见挖掘/0.008
网络谣言分析	网络舆情/0.642、网络谣言/0.103、传播模式/0.037、情感分析/0.035、电子政务/0.025、舆情监控/0.019、沉默的螺旋/0.017、案例分析/0.014、舆情管理/0.012、公众情绪/0.010

表9 新闻传播学科的 LDA 主题模型挖掘结果

主题标签	前 10 个重要主题词及权重
基于传播学的舆情分析	网络/0.702、社会网络/0.023、躲猫猫/0.021、社会政治态度/0.019、传播学研究/0.019、网络新闻/0.014、思想文化信息/0.012、新闻线索/0.012、新华网/0.010、天涯社区/0.010
网络舆情监测	人民网/0.139、舆情分析/0.139、舆情/0.131、舆论监督/0.106、舆情监测/0.087、舆情研究/0.049、大数据/0.032、监测室/0.028、监测/0.025、传播主体/0.021
网络舆情传播	新媒体/0.445、舆论/0.327、传播模式/0.025、媒体融合/0.012、社会化媒体/0.012、躲猫猫/0.010、网络推手/0.008、舆论监督/0.008、媒体环境/0.008、网络论坛/0.006
舆情危机处理	网络舆情/0.649、舆情危机/0.21、民意表达/0.020、线下/0.020、传播能力/0.018、微博/0.016、网络民意/0.014、公众情绪/0.012、社交媒体/0.012、沉默的螺旋/0.012
突发事件舆情传播	舆情/0.360、突发事件/0.279、网络舆情/0.105、舆论/0.047、网络民意/0.022、高校/0.020、网络媒体/0.017、沉默的螺旋/0.015、食品安全/0.014、传播渠道/0.012
网络问政	网络问政/0.402、政府/0.317、信息传播/0.038、媒介素养/0.034、把关人/0.020、政府公信力/0.015、陈力丹/0.013、国家管理者/0.012、舆情事件/0.012、社会空间/0.010
大数据舆情传播	互联网/0.331、信息传播/0.056、新媒体/0.050、大数据/0.036、网络暴力/0.035、群体极化/0.033、广播电视/0.033、网络事件/0.031、传播力/0.031、舆情/0.029
意见领袖	意见领袖/0.184、传统媒体/0.152、议程设置/0.093、自媒体/0.072、热点事件/0.065、网络新闻/0.040、网络谣言/0.037、地方政府/0.033、舆论场/0.026、引导机制/0.023
高校舆情分析	网络舆情/0.550、高校/0.105、群体性事件/0.090、预警/0.021、思想政治教育/0.021、公众情绪/0.017、意见领袖/0.017、中国青年报/0.012、社会化媒体/0.012、舆论导向/0.009
网络舆论引导	网络舆论/0.374、舆论引导/0.216、舆论场/0.069、网络谣言/0.027、网络论坛/0.022、网络编辑/0.022、突发事件/0.020、新闻发言人/0.017、舆论传播/0.013、网络公关/0.011

综上分析可得出，网络舆情在图书情报学领域的热点主题有：研究热点分析、节点影响力分析、基于情报学的舆情分析、高校舆情分析、政府舆情应对、网络舆情预警与监测、大数据舆情分析、网络舆情演化、网络谣言分析、舆情话题发现、危机处理、突发事件舆论引导等；新闻传播学领域的网

络舆情研究热点有：基于传播学的舆情分析、网络舆情监测、舆情危机处理、突发事件舆情传播、高校舆情分析、网络舆论引导、大数据舆情传播、意见领袖分析、议程设置等。

另外，从表8和表9所示的结果中还可以发现，两门学科除了具有交叉主题外，还存在着一些具有特色化的主题，例如，涉及大数据技术处理的网络舆情及其演化分析、网络谣言分析、基于情报学的舆情分析等是图书情报学领域的特色主题，而涉及内容传播与民意表达的网络问政和基于传播学的舆情分析则是新闻传播学领域的特色主题。

### 3.5 讨论

上述研究结果表明，两门学科在网络舆情研究方面既存在共性和差异性，又衍生了新兴的交叉主题增长点。

(1) 两门学科的部分研究主题相同，但两者的关注视角有所差异。综合基于共词聚类的划分结果和基于LDA的挖掘结果可以看出，两个学科都重点关注了突发事件/热点事件的舆情传播与监测、高校舆情分析、大数据舆情分析、危机预警与管理、政府舆情应对等主题，但两者对网络舆情的关注视角有所差异，图书情报学主要关注舆情，而新闻传播学侧重舆论研究。此外，两者共同关注的主题在两门学科中的发展态势也不一样：舆情监测与预警、突发事件舆论引导两个主题在两门学科中都发展得较为成熟；节点影响力分析和政府舆情应对是图书情报学中的核心主题；舆情危机处理和高校舆情分析是新闻传播学中的核心主题。然而，其他主题则在两门学科领域中可能处于起步阶段，发展尚不稳定。

(2) 两门学科不但有共同关注的主题，而且相同主题所限定的范围也基本一致，但其涉及的内容不同。网络舆情在发展之初，其关注范围就被限定到Web2.0环境下的用户生成内容，包括BBS、网络论坛、网络社群、门户网站等。但随着自媒体时代的到来，两门学科的学者逐渐将研究重点转向以微博为代表的新型社交媒体上，同时也将关注焦点转移到社会突发事件或热点公共事件的舆情传播与监测。虽然两门学科在部分主题的限定范围上具有相似之处，但是两者所涉及的内容有所差别，图书情报学多从技术角度研究突发事件的信息处理、舆情监测与演化分析、网络谣言分析、事件传播规律与行为分析等，而新闻传播学则主要从内容角度研究突发事件的民意监测、议程设置和舆论引导等。

(3) 两门学科涉及的知识点类似，但采用的理论方法有所区别。两门学科在网络舆情研究方面都共同关注了意见领袖、舆情监测、公共危机、舆情传播、舆情引导等知识点，但两者所采用的理论方法不同。图书情报学领域的网络舆情研究大多是基于网络信息内容的挖掘与利用，研究重点集中在网络信息内容之间的社会关系网络挖掘、网络信息流在网民之间传播扩散的动力学行为挖掘、网络舆情的社会影响评估等，偏向于理论方法和信息处理技术的研究，特别是较多地引入了自然语言处理、机器学习和数据挖掘等交叉领域的信息处理技术，如分类学习、聚类学习、内容分析法、共词分析法、情感分析、LDA、层次分析法、社会网络分析法和神经网络分析法。新闻传播学多从传播学视角以定性描述方法为主，重点分析网络舆情传播、舆论引导、议程设置、舆情危机应对、谣言传播、网络问政等内容的重要性、特点和策略，但对于网络舆情的传播形成机理、发展与动态演化过程，缺乏深刻的科学认知，且对于地理空间原因、网民情感倾向对舆情传播的内在作用机制缺乏理性认识。

(4) 学科交叉融合孕育新的潜在交叉主题和衍生主题。从研究的广度和深度来看，两门学科对于网络舆情的研究越来越关注网民之间的社会关系、情感倾向、信息传播模式等特征，而对于网络社交媒体信息交互的可视化、虚假信息传播、非理性网民情绪、意见领袖的角色功能与引导等研究方向缺乏足够的认识和分析。但是，当前中国网络舆情事件频发，社会舆情事件与涉腐涉政舆情事件不断涌现，使得网络舆情成为了政府开展突发事件应急管理的重要辅助决策情报源。在这种背景下，两门学科的学者逐渐将研究主题聚焦到大数据环境下突发事件的知识图谱分析、意见领袖识别与引导、网络谣言检测和社会公共危机管理等具有较强交叉性的主题上，从而孕育出了更多新的潜在交叉主题和衍生主题。

## 4 结论

本研究在关键词规范化、数据预处理、关键词共现矩阵构建等处理流程的基础上，以关键词交集分析、共词聚类、战略坐标分析以及LDA主题模型为手段，分析了两学科在网络舆情研究方面的热点主题及其分布。研究表明，两门学科在网络舆情的研究内容和主题分布方面既存在着共性，又存在着特性。当然，由于学科的相关性和交叉性特点，两

学科在该研究领域上也衍生出了一些新的交叉主题增长点。虽然在研究方法上,通过将共词聚类 and LDA 主题建模进行相结合能够在一定程度上弥补彼此的缺陷,但是忽略了关键词的时序变化特性以及主题的动态演化过程。

尽管近两年的国内网络舆情研究在发展速度上稍有减缓,但是从国家宏观战略上看,该研究将会在“十三五”期间再次出现一个新的发展机遇。主要原因如下:一是,国家在宏观层面上给出了更多的政策支持,例如,国家在“十三五”规划中明确提出了“创新符合网络传播规律的网上宣传方式和提升网络舆情分析和引导能力”的网络文化建设目标;二是,随着网络空间安全和“一带一路”战略的提出,大数据、云计算、深度学习、自然语言处理、数据挖掘、人工智能以及机器学习等微观层面上的前沿技术必将得到快速发展,这也将会促使网络舆情研究在内容、理论、技术上出现新的突破和创新。

### 参 考 文 献

- [1] 李昌祖,许天雷.网络舆情与网络舆论的辩证关系[J].浙江工业大学学报(社会科学版),2009,8(4):393-398.
- [2] 曾润喜.网络舆情信息资源共享研究[J].情报杂志,2009,28(8):187-191.
- [3] 高承实,陈越,荣星,等.网络舆情几个基本问题的探讨[J].情报杂志,2011,30(11):52-56.
- [4] 季丹,谢耘耘.中国网络舆情研究的历史回顾与反思——基于CNKI、CSSCI高被引论文观察[J].上海交通大学学报(哲学社会科学版),2012,20(4):48-56.
- [5] Wang G H, Liu Y J, Li J M, et al. Superedge Coupling Algorithm and Its Application in Coupling Mechanism Analysis of Online Public Opinion Supernetwork[J]. Expert Systems with Applications, 2015, 42(5): 2808-2823.
- [6] Liu H. Internet Public Opinion Hotspot Detection and Analysis Based on K-means and SVM Algorithm[C]. International Conference of Information Science and Management Engineering, 2010: 257-261.
- [7] 谢耘耘.舆情蓝皮书:中国社会舆情与危机管理报告[M].北京:社会科学文献出版社,2015.
- [8] 毕宏音.现代舆情研究十年历程的回顾和反思[J].天津社会科学,2013(4):67-71.
- [9] 李长玲,纪雪梅,支岭.基于E-I指数的学科交叉程度分析——以情报学等5个学科为例[J].图书情报工作,2011,55(16):33-36.
- [10] 李长玲,郭凤娇,支岭.基于SNA的学科交叉研究主题分析——以情报学与计算机科学为例[J].情报科学,2014,32(12):61-64.
- [11] 张艺蔓,李秀霞,韩牧哲.基于引文耦合的情报学学科结构时序分析[J].情报杂志,2015,34(3):100-106.
- [12] 闵超,孙建军.学科交叉研究热点聚类分析——以国内图书情报学和新闻传播学为例[J].图书情报工作,2014,58(1):109-116.
- [13] 钱玲飞,杨建林,张莉.基于关键词分析的学科创新力比较——以情报学图书馆学为例[J].情报理论与实践,2011,34(11):117-120.
- [14] 胡昌平,陈果.科技论文关键词特征及其对共词分析的影响[J].情报学报,2014,33(1):23-32.
- [15] 王连喜,李霞.国内微博研究热点分析及主题挖掘——以计算机和图书情报学为研究对象[J].情报杂志,2015,34(4):127-132.
- [16] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet Allocation[J]. Journal of Machine Learning Research, 2003(3): 993-1022.
- [17] 丁玉飞,王曰芬,刘卫江.基于主题模型的科技监测方法及应用研究[J].情报学报,2015,34(8):854-865.
- [18] 崔雷,隋明爽.共现聚类分析结果表达方法的研究[J].情报学报,2015,34(12):1270-1277.

(责任编辑 魏瑞斌)