信息管理
新视野论丛

JI YU WEN BEN TE ZHENG JI SUAN DE

基于文本特征计算的信息分析方法

XIN XI FEN XI FANG FA

许 鑫 著

上海科学技术文献出版社
Shanghai Scientific and Technological Literature Press

JI YU WEN BEN TE ZHENG JI SUAN DE

基于文本特征计算的信息分析方法

XIN XI FEN XI FANG FA

许 鑫 著

上海科学技术文献出版社
Shanghai Scientific and Technological Literature Press

图书在版编目（CIP）数据

基于文本特征计算的信息分析方法 / 许鑫著．—上海：上海科学技术文献出版社，2015.11

ISBN 978-7-5439-6835-6

Ⅰ．① 基…　Ⅱ．①许…　Ⅲ．①信息—分析方法　Ⅳ．① G202

中国版本图书馆 CIP 数据核字 (2015) 第 227466 号

责任编辑：徐　静

封面设计：徐　炜

基于文本特征计算的信息分析方法

许　鑫　著

出版发行：上海科学技术文献出版社

地　　址：上海市长乐路 746 号

邮政编码：200040

经　　销：全国新华书店

印　　刷：上海市印刷七厂有限公司

开　　本：787×1092　1/16

印　　张：15.5

字　　数：377 000

版　　次：2015 年 11 月第 1 版　2015 年 11 月第 1 次印刷

书　　号：ISBN 978-7-5439-6835-6

定　　价：45.00 元

http://www.sstlp.com

目　录

绪　论

信息分析是一种以信息为研究对象，根据拟解决的特定问题的需要，收集与之有关的信息进行分析研究，旨在得出有助于解决问题的新的信息的科学劳动过程。面对大数据环境下的今天，要寻求行之有效的方法与之相适应，这样才能从大量无序数据中提炼出有价值的信息。信息分析方法的优劣直接关系到信息分析结果的好坏；当然，简单武断地说某种信息分析方法好或者不好并不是科学严谨的态度，用适合或者不适合可能是更加合适的表述。

信息分析要以社会用户的特定需求为依托，通过对大量信息进行收集和整序，消除其中的不确定性因素，以使信息增值。这一过程中，涉及信息的收集、整理、鉴别、评价、综合等系列化加工过程，可以说它是一项实用性很强的科学，信息分析方法是在实践中不断积累和发展而形成的方法体系。既然是科学方法体系，它就不是一成不变的，而是与研究对象和研究领域的发展变化相一致的，随着研究领域的拓展而不断丰富和充实。方法在其发展中呈现它自身的规律性。科学方法的应用离不开科学理论的指导，尤其在社会信息化的今天，对于信息分析方法论的探讨显得更加重要。

方法论并不等同于方法。方法论是对方法进行研究的科学，是比方法更高层次的东西。方法论是对各种方法的优劣、关系、功能进行评价、综合的研究，抽象出一些共有的东西，使之上升到更高的层次。方法论是科学的一个分支，它关注的是科学调查的方法和技巧，特别关注调查特定技巧或程序的潜力和局限性。一个特定的方法论的途径将会由具体的本体论和认识论的假设所支撑，而且这个方法论的途径也会反映出具体的本体论和认识论的假设，这些假设强调理解和认识这个世界的方法，在特定研究中选定某种途径和方法。方法论解决的是调查的逻辑以及如何形成理论和随后该如何验证理论等问题[1]。

我国的传统信息分析研究起步于20世纪中期，走的是文献工作和研究工作相结合的道路。初期阶段主要是以文献工作为主，基本任务是摸清世界各国科学技术的先进水平，重点配合国家科学发展规划与经济发展规划的制定、科研与生产攻关项目的实施，以及提高科研、仿制和生产的能力，全面、及时、准确地反映国外科技发展的新水平、新动向。十一届三中全会之后，信息分析研究工作面向经济建设主战场，大大拓宽了服务范围，逐步提升到以综合研究和决策研究为主，从事大范围、跨行业、多学科、多因素的信息分析研究，为科学决策提供依据。我国的信息分析研究是在20世纪80年代以后大量吸收国外的新观念、新技术、新方法而迅速发展起来的，但我国自己的理论体系却没有很好地建立起来[2]。这其中

---

〔1〕[英]乔纳森·格里斯. 研究方法的第一本书[M]. 孙冰洁，王亮，译. 大连：东北财经大学出版社，2011.

〔2〕张帆. 关于信息分析方法论研究的几点看法[J]. 图书馆学研究，2000(1)：38—39.

可能存在两方面的问题。一方面，方法的吸收和移植是信息分析方法论研究的重要问题，但信息分析本身缺乏自己特有的方法。从学科归属角度来看，它应该与信息科学密切相关，甚至可以把信息分析和情报分析作为姊妹概念，但图书情报领域特有的一些理论和方法在其他学科研究以及具体实务领域应用得并不广泛，而信息分析方法中的一般方法——如社会调查法、数学方法等——则是来源于其他相关学科的研究成果和方法的移植。另一方面，现有的信息分析研究更多地侧重于方法的具体应用而缺少必要的归纳总结，更谈不上更高层次的理论探讨了。目前方法论研究的薄弱已不能满足信息分析的发展，理论研究跟不上实践的需要，势必会阻碍信息分析研究的改进和创新。

互联网的蓬勃发展使得越来越多的信息分析基于网络信息资源来开展，海量的数据使得用人工进行数据处理和分析变得越来越困难，尤其是大数据时代的到来，越来越多的用于分析的数据呈现出海量、多源、非结构化等特点。这客观要求我们必须改进原来的方法以适应时代需要，尤其应该在具体应用各种方法时避免定性与定量分离，充分利用现代计算机技术对客观事实、数据进行分析。文本数据是常见的一类数据形式，不管是数据库中的文本数据类型，还是各类电子化的文档资源，抑或是 Web 文本，甚至图片、音频、视频等资源的标签，在不同的场合下都可能是我们进行信息分析的数据源。文本中蕴含着大量的概念和知识，也是实时监测中各类信号和线索的载体，对其有效的处理和正确的解读是信息分析的基础，也是情报分析中不可或缺的一类研究；同时，信息分析方法论的完善关系到信息分析方法能否更好地利用相关学科的优秀成果，并将自己的特有方法向其他学科转化。

在本书中，我们不敢奢谈方法论研究，仅仅希望聚焦文本类型的数据，通过对文本特征的提取、计算和分析，融合定性研究和定量研究形成统一的分析框架，并结合领域实例归纳总结出一套行之有效的信息分析通用方案，能够对多领域的研究工作开展和社会经济生活中的分析实务提供支撑。

第1章

信息分析方法概述

1.1 定性分析与定量分析

一般而言，信息研究的方法可以划分为定性研究、定量研究以及定性与定量研究方法的组合。定性研究侧重于用语言文字描述、阐述以及探索事件、现象和问题；定量研究侧重于用数字来描述、阐述以及揭示事件、现象和问题；定性与定量研究方法的组合则是定性与定量两种方法兼而有之。需要指出的是，具体问题的研究中，严格划分研究方法有时是较为困难的，因为定性研究不等于没有数字，而定量研究中也不乏直觉、价值判断和逻辑推理等，因此定性研究与定量研究在不少情形中往往是融合交杂在一起的。

1.1.1 定性研究方法

定性研究是一组跨学科、跨专业、跨领域、跨主题的研究方法，由一组复杂的、相互关联的术语、概念和假设等组成[1]。通过陈列定性研究的具体方法及其适用的研究领域，可以更好地理解什么是定性研究。定性研究方法主要包括：人种学；参与观察；文化人类学；文件、经书、符号和叙述分析；案例研究；档案分析；内容分析；通信分析；符号的相互作用分析；种族方法论；心理分析；女性主义追问；现象学；问卷研究；结构解剖；行动研究和参与式行动研究；访问研究；后实证主义研究；后结构主义研究等[2、3]。

定性研究通常意味着三个概念：(1)构建的研究认识论(即基于认识知识的后现代、结构主义或自然主义范式的方法)；(2)具体的研究战略，如研究设计是针对解释和揭示事物、现象和事件，而不是总结出可运用于更大范围的因果关系；(3)具体的、不需要涉及数字的技术，如访问法。简言之，定性研究方法是由访问、观察、案例研究等多种方法组成。原始资料包括场地笔记、访谈记录、对话、照片、录音和备忘录等，目的在于描述、解释事物、事件、现象、人物并更好地理解所研究问题的研究方法。

学界对定性研究的优劣及可靠程度的检验缺乏共识，不过大多数研究都遵循以下四项标准：内部有效性；外部有效性；可靠性；客观性。Miles 和 Huberman 提出下列标准确定定性研究程序的优劣：(1)在案例内或案例之间的抽样决策；(2)数据收集操作；(3)数据库的规

[1] DENZIN N K, LINCOLN Y. The SAGE Handbook of Qualitative Research [M]. Sage, 2011.
[2] MARSHALL C, ROSSMAN G B. Designing Qualitative Research [M]. Thousand Oaks, 1995.
[3] DENZIN N K, LINCOLN Y. Qualitative Research [J]. Thousand Oaks, 2000.

模和总结，产生数据的方法；(4)软件的使用；(5)总体的分析战略；(6)关键数据的展示以支持主要结论[4]。Strauss 和 Corbin 提出判断定性研究的四项标准：(1)资料的有效性、可靠性和信誉度；(2)对理论自身的判断；(3)产生、阐释、测试理论的研究过程是否完善；(4)结论的得出是否扎根于经验的数据[5]。Marshall 和 Rossman 陈述了 20 个问题以判断定性研究的质量[6]，限于篇幅，不再赘述。

虽然学界提出了各种各样的定性研究的标准，但在我国的实际研究中对定性研究还是存在着认识上的误区。因为定量研究对研究者的数学、统计学以及计算机知识都有相当的要求，因此一些缺乏或缺少定量研究方面训练的研究者更倾向于采用定性研究方法。由于有不少学者缺乏社会科学研究方法的训练，他们简单地把不能对所收集的资料做量化分析，只能以文本的形式进行描述的研究都看作是“定性研究”，既不愿意花费比较长的时间对研究对象进行长期深入的观察和体验，也不愿意充分发挥自主性和主动性与研究对象进行互动，深入地理解研究对象行为及其环境的真实意义，更多的是对一些二手资料进行加工处理。这样一来，使得我们的很多研究方法不够规范，研究质量也存在着一定的问题。

1.1.2 定量研究方法

定量化研究是针对可计量研究对象，利用某些定量化研究资源，采用一定的定量研究方法和手段，寻求定量表示以揭示数据之间的关系和规律的一种研究方法。以情报学中的定量研究为例，它最初是从文献计量学开始，随着情报研究活动的深入开展，定量化研究逐渐延伸到网络信息等其他领域，派生出信息计量学、网络计量学等其他的分支。由于现代科技、经济和社会的不断发展，信息资源的网络化及科学研究的信息化和社会化，使得情报学处于急剧变革和发展之中，也更直接地影响到了情报研究工作，其中定量研究一直是情报学乃至整个科学发展的重要方向和必然趋势。正如英国著名情报学家布鲁克斯所指出的：“情报学如果不实现定量化，它将是一堆支离破碎的技艺，而不会成为科学。”因此实现情报研究定量化，才能提高情报学的科学性和精确性，从而有助于确立和提高情报学在整个科学体系中的学科地位。

信息分析作为情报学研究范畴中的重要领域，通过定量分析可以用精确的数量值代替模糊的印象，可以依据数学公式导出精确的数量结论，还可以将结论的数量形式解释为直观性质，其特点在于可以对事物的发展做出定量的描述，借助于数和数量关系研究事物的发展规律[7]。但定量的信息分析研究也有其不足之处：首先，定量研究方法是在获取目标样本过去或现在信息的基础上进行研究的，无法对目标样本的未来、瞬间状态进行跟踪、表达和研究；其次，定量研究方法力求获得精确的结论，虽然承认“偏差”的存在但未必引起足够的重视，结果可能获得违反生活常识的结论；再次，信息分析人员有可能陷入“数据的精确性、结论的严密性”的误圈，被“客观”的数据和他无法“精确”解释的某些现象所迷惑，最终影响决策的顺利完成[8]。

---

[4] MILES M B, Huberman A M. Qualitative Data Analysis: An Expanded Sourcebook [M]. Sage, 1994.

[5] STRAUSS A, CORBIN J. Basics of Qualitative Research [M]. Thousand Oaks, 1998.

[6] MARSHALL C, ROSSMAN G B. Designing Qualitative Research [M]. Thousand Oaks, 1995.

[7] 丁建琴. 情报探索[M]. 无锡：江南大学出版社，2011.

[8] 王崇德. 文献计量学引论[M]. 桂林：广西师范大学出版社，1997.

1.1.3　定性与定量相结合

从上文对定性研究和定量研究的简单介绍和分析中可见，定性研究包括了定性的比较、分类、类比、分析和综合、归纳和演绎，可以获得研究对象质的特性，而定量研究要获得关于研究对象量的特征，包括各种测量方法、定量实验方法和数学方法。传统的信息分析研究多用定性分析方法，如信息报道、专题文献索引等，但是这种靠经验总结来进行的分析活动很难保证分析结果的准确性和重复性，也很难使信息分析活动成为一种严格意义上的科学研究活动。定量分析方法强调对数据的分析，通过建立数学模型等可重复检验的手段表达数据的内涵，这样才能保证信息分析活动成为建立在可靠基础上的科学活动。

随着时代——特别是科技和经济——的发展，信息分析研究的主要对象从科学技术领域转移到了经济领域，调查结果表明政府和企业是信息分析的主要服务对象，而决策研究决定了信息分析工作的主流方向。市场经济体制下，市场需求决定了信息分析必须借助于先进的方法和手段，将定性分析和定量分析结合起来使用，才能满足决策、预测的需要[9]。计算机和网络的迅猛发展使人们可以获得更多、更及时的信息，软件技术的发展使得对大量数据进行反复的运算成为可能，这为定量分析的发展带来了极大的方便。

定性研究与定量研究的有机结合改变了传统信息分析方法多用定性研究方法、很难保证分析结果准确性和重复性的局限。一方面，定性研究把握信息研究的重心和方向，侧重于物理模型的建立和数据意义；另一方面，定量研究为信息分析结果提供数量依据，侧重于数学模型的建立和求解。通过定性研究与定量研究方法的结合，使得信息分析方法更加符合实际需要，得出的结论更加准确和可靠。定性分析是定量分析的基础和前提，定量分析则是定性分析的精确和具体化，两者对立统一。在实际工作中，定性分析的终点常常是定量分析的起点。信息分析的理论与实践证明，传统的定性分析方法与现代的定量分析方法的有机结合，将是信息分析方法的必然走向。定量分析和定性分析是信息分析的两个方面，二者缺一不可。

科技发展促进了学科之间的交叉渗透，也推动了信息分析方法向多元化方向发展，因此对于信息分析方法体系的研究也呈现出多元化的趋势。虽然信息分析本身比较缺乏特有的方法，但它吸收、移植、借鉴和综合了其他学科的方法，不管是定性研究方法，还是定量研究方法，在信息分析中对其具体应用形式加以研究，并与信息分析的实际需求结合起来，必将能形成自身的特色，进而有可能进一步创新发展，并最终形成特有的信息分析方法论。

1.2　常用信息分析方法

信息分析是以定性和定量研究方法为手段的，所以我们关注信息分析方法首先就要关注与之相关的一些定性研究方法、定量研究方法，或者两者的结合。除了对方法的掌握，信息分析也需要从成因、过程、成果、目的等角度深入理解，只有这样才可能形成更好的使用方法。一般说来，信息分析的产生是由于存在社会需求；过程一般会涉及课题选择、问题界定、

[9] 查先进. 信息分析与预测[M]. 武汉：武汉大学出版社，2000.

信息收集、信息整理、分析与预测、信息产品制作与利用等流程；成果主要体现在新的增值的信息产品上；而目的是为不同层次的科学决策服务[10]。信息分析涉及的方法很多，目前信息分析的方法主要是从图书情报、管理学、技术经济、数学、统计学、计算机等学科领域借鉴或移植过来的，其中使用较为广泛的是一些情报学和软科学的研究方法。下文将介绍与本书相关的一些信息分析方法。

1. 逻辑思维法

信息分析中的逻辑思维方法是建立在逻辑推理和辩证分析基础上，根据已知信息运用分析和综合、演绎与归纳、相关与比较等一系列逻辑思维手段来揭示研究对象的本质、发展规律和因果关系，具有广泛分析、定性使用、推理严密的特点；不过，它虽然具有很强的说理性但是并不具体，推理虽严密但不够精确，其结果往往缺乏定量的表述和结论，仅仅是一种定性认识或描述，所以它不太适合一些需要定量化研究的信息分析课题。

在诸多具体的逻辑思维方法中，比较法是用得比较多的一类分析方法。比较法实际上就是对研究对象的某些共同特性或属性进行对比，所以在对比时必须对反映事物本质的特征或属性进行分解和分析，并从中确定并分析其特征和属性。通过比较，可以发现事物间本质上的异同，揭示国家、地区、行业、部门、产品、技术、工艺等当前的水平和差距，以便于对比发展水平，明确发展方向；同样，通过对事物不同时期发展状况和水平的比较，也可以了解其发展轨迹，揭示其发展规律，判明其发展方向，以便于认识事物的过去和预测事物的未来；还可以通过比较不同的方案明确优劣、真伪，从而为识别、判断和选择提供依据。

其他的逻辑思维方法还包括分析、综合、推理等。分析是把客观事物整体分解为部分或要素，并根据事物之间或事物内容各要素之间的特定关系，通过推理、判断达到认识事物的目的；而综合正好与之相对立，人们可以将研究对象片面、分散、众多的各个要素（情况、数据、素材等）进行归纳，从错综复杂的现象中探索它们之间的相互关系，从整体上把握事物的本质和规律，通观事物发展的全貌和全过程；至于推理，则是从一个或几个已知的判断得出一个新判断的过程，具体而言就是在掌握一定的已知事实、数据或因素相关性的基础上，通过因果关系或其他相关关系顺次、逐步地推论，最终得出新结论的一种方法。

2. 专家调查法

在信息分析方法中，定性与定量相结合的研究方法的典型代表就包括了德尔菲法（又称为专家调查法），它是在20世纪40年代由赫尔姆和达尔克首创，以古希腊城市德尔菲命名的。1946年美国兰德公司为避免集体讨论存在的屈从于权威或盲目服从多数的缺陷，首次用这种方法进行定性预测，后来该方法被迅速广泛采用。该方法的特点是简便直观，无须建立繁琐的数学模型，能够较精确地反映出专家们的主观判断力，是目前从事未来分析预测时常使用的一种方法。除了用于科技领域预测外，它还被用于其他领域的预测，如军事预测、人口预测、医疗保健预测、教育预测等，大多取得了良好效果，同时它还被用来进行评价、决策和规划工作，在长远规划者和决策者心目中享有很高的可信度。

德尔菲法采用背对背的通信方式征询专家小组成员的意见，经过几轮征询，使专家小组的意见趋于集中。依据一般的原则和程序，德尔菲法应用过程中采用匿名发表意见的方式，

[10] 苏敏. 信息分析方法在信息素质教育中的应用研究[J]. 大学图书情报学刊，2010(3)：60—63.

即团队成员之间不得互相讨论，不发生横向联系，只能与调查人员发生关系。反复填写问卷，以集结问卷填写人的共识及收集各方意见，可用来构造团队沟通流程。利用德尔菲法进行信息分析的程序通常包括：确定研究课题并制定实施计划；选择和组织专家；设计调查表，实施多次反馈；对最后的调查结果进行分析和数据处理以及评价和预测。

德尔菲法经过半个多世纪的发展，已经有很多派生的方法产生，如加表德尔菲法、加测德尔菲法、加评德尔菲法、加因德尔菲法以及头脑风暴法等。其中值得一提的是头脑风暴法，在信息分析中也广受欢迎。它强调的几条原则包括：禁止批评他人的建议，只许完善；最狂妄的想象是最受欢迎的；重量不重质，任何一种构想都可以被接纳；禁止参与者私下交流，以免打断他人的思维。使用头脑风暴法可以有益于激发参与者更多的创新思维和得到意想不到的解决方法。

3. *层次分析法*

层次分析法是美国著名运筹学家、匹兹堡大学 T. L. Saaty 教授于 20 世纪 70 年代中期提出的。层次分析法由于在解决多目标决策问题方面具有比其他方法更简便实用的特点，因而被广泛应用在政治、经济、社会等领域。层次分析法的整个过程体现了人的决策思维活动中的分析、判断、综合等基本特征，因此也被看做是定性分析与定量分析较好结合的信息分析方法。层次分析法作为系统性的分析方法简洁实用，但它并不能为决策提供新方案，且在指标过多时数据统计量大且权重难以确定。

层次分析法的基本步骤是：(1)将问题概念化，找出研究对象所涉及的主要因素；(2)分析各因素的关联、隶属关系，构造系统的递阶层次结构；(3)对同一层次的各因素关于上一层次中某一准则的重要性进行两两比较，构造判断矩阵；(4)由判断矩阵计算被比较因素对上一层次该准则的相对权重，并进行一致性检验；(5)计算各层次因素对于最高层次，即系统目标的合成权重，进行层次总排序，并进行一致性检验。

4. *回归分析法*

回归一词最早来源于英国生物学家兼统计学家 Galton 对遗传现象的大量观察统计。回归分析法是通过研究两个或两个以上变量之间的相关关系来对未来进行分析与预测的一种数学方法，它不仅提供了建立变量之间相关关系的数学表达式的一般途径，而且可以对所建立的经验公式的适用性进行分析，广泛地应用于分析预测以及控制等方面。

回归分析的步骤大致如下：(1)根据自变量与因变量的现有数据以及关系，初步设定回归方程；(2)求出合理的回归系数，确定回归方程；(3)进行相关性检验，确定相关系数；(4)在符合相关性要求后，可确定事物的未来状况，并计算预测值的置信区间。

回归分析通过处理已知数据来探寻这些数据的变化规律，并以此建立相应的回归方程式，再根据该方程式来预测未来发展的一种数理统计方法，它具体可以分为一元、二元和多元线性及非线性回归法等[11]。

回归分析法在分析多因素模型时比较简单和方便，而且运用回归模型，只要采用的模型和数据相同，通过标准的统计方法就可以计算出唯一的结果——当然，在图和表的形式中，数据之间关系的解释往往因人而异，不同分析者画出的拟合曲线很可能也是不一样的；同

---

[11] 郭吉安，李学静. 情报研究与创新[M]. 北京：科学出版社，2006.

时，回归分析可以准确地计量各个因素之间的相关程度与回归拟合程度的高低，以提高预测方程式的效果。回归分析的局限性主要体现在选用何种因子和该因子采用何种表达式只是一种推测，这影响了使用因子的多样性和某些因子的不可测性，从而使得回归分析在某些情况下受到限制。

5. 时间序列法

时间序列分析法实际上是一种特殊的回归分析法，它不再考虑事物之间的因果关系或其他相关关系，而仅考虑研究对象与时间之间的相关关系，即将时间作为自变量。时间序列分析预测就是将时间序列上构成波动的不同数据类型分离开来，分别进行分析，找出事物随时间的变动规律，并以此为依据预测事物的未来状况。

时间序列分析是定量预测方法之一。它的基本原理一是承认事物发展的延续性，应用过去的数据，就能推测事物的发展趋势；二是考虑到事物发展的随机性，任何事物的发展都可能受偶然因素影响，为此要利用统计分析中加权平均法对历史数据进行处理。该方法简单易行，便于掌握，但准确性差，一般只适用于短期预测。时间序列法又分为倾向线拟合法和倾向线逐步修正法，前者包括多项式曲线、指数曲线和生长曲线，后者包括移动平均法和指数平滑法。

6. 文献计量法

基于文献量的变化与科学技术的发展之间存在着一定的内在联系，从而可以利用文献量的变化建立表征这一内在联系的方程式，并据以了解科学技术的历史、现状和发展趋势。文献计量用数学和统计学的方法，定量地分析一切知识载体，其计量对象主要包括文献量（各种出版物，尤以期刊论文和引文居多）、作者数（个人集体或团体）、词汇数（各种文献标识，其中以叙词居多）等。文献计量学是集数学、统计学、文献学为一体的交叉科学，它注重量化综合性知识体系，所以其最本质的特征在于其输出务必是“量”。

文献计量是以几个经验统计规律为核心的，例如，表征出科技文献作者分布的洛特卡定律（1926），表征文献中词频分布的齐普夫定律（1948），确定某一学科论文在期刊中分布的布拉德福定律（1934）等。围绕这几个定律，现在基于此类方法的信息分析应用很广泛：微观的应用可以确定核心文献，评价出版物，考察文献利用率，实现图书情报部门的科学管理；宏观应用可以辅助设计出更经济的情报系统和网络，提高情报处理效率，寻找文献服务中的弊端与缺陷，预测出版方向等。

因为存在着影响文献情报流的人为因素，所以很多文献问题尚难以定量化，特别是由于文献系统高度的复杂性和不稳定性，我们不可能获得足够的、有效的信息来揭示文献的宏观规律，所以文献计量方法的发展也有赖于数学工具和统计学技术的进一步支持；不过，它在拓展文献计量应用研究对象方面还是有着广阔的空间，信息计量、网络计量、科学计量等也都成长为基于不同载体的信息分析方法。

7. 内容分析法

内容分析法（content analysis）是一种对研究对象的内容进行深入分析，透过现象看本质的科学方法，是一种基于定性研究的量化分析方法。它将用语言表示而非数量表示的文献转换为用数量表示的资料，并将分析的结果用统计数字来加以描述。内容分析法具有三个基本要素，即客观、系统、量化。其原理是对文献内容所含信息量及其变化进行分

析，从而对文献内容进行可再现的、有效的推断[12]。内容分析法的过程较为复杂，其基本步骤为：提出研究问题、抽取文献样本、确定分析单元、制订目类系统、内容编码与统计、解释与检验。

内容分析法也是一种定量和定性相结合的研究方法，纯定性的研究已难以满足科学研究的需求。任何事物都是质和量的统一，把事物的量作为一种测量工具，对事物的质进行精确量化，不但可信而且有利于对质的系统研究和了解[13]。内容分析法作为一种研究社会现实的科学方法，经历了不断的理论探讨和实践应用，正逐步趋于成熟与完善。内容分析法作为一种信息分析方法虽然具有客观、系统、量化的特点，也克服了其他纯定性或纯定量分析方法的缺陷，但它本身也存在一定的局限性，主要表现为：(1)抽样调查的人为因素。内容分析法虽然是基于大量的文献分析，但揽括全面是不可能的，必须进行一定的人为抽样，这个过程存在一定的主观判断。(2)手工标引的低效率。内容分析能够将非结构化的媒体内容转换为量化和结构化的内容进行分析，这个过程是通过编码标引来完成的。目前这个阶段可以通过计算机辅助查找，但核心的标引过程还是基于手工操作。这就不可避免地带来一个效率问题。(3)长期作业或集体作业的信度问题。内容分析是一个复杂且繁琐的过程，常常需要少量工作人员长时间的操作或多人的合作，这又不可避免研究信度问题[14]。因为在标引过程中，无论一个人对一批样本长时间进行标引还是多人对一批样本同时标引，都可能导致标引的不一致或标引的信度不高的问题。

8. 社会网络分析

社会网络分析(social network analysis，SNA)方法是由社会学家根据数学方法、图论等发展起来的定量分析方法。社会网络分析是社会学领域比较成熟的分析方法，社会学家们利用其可以较为得心应手地解释一些社会学问题。许多学科，如经济学、管理学等领域的学者们在新经济时代——知识经济时代面临许多挑战时，开始考虑借鉴其他学科的研究方法，社会网络分析就是其中的一种。近几年来，该方法已经在职业流动、城市化对个体幸福的影响、世界政治和经济体系、国际贸易等领域的分析中发挥了重要作用。

网络指的是各种关联，而社会网络(Social Network)即可简单地称为社会关系所构成的结构。社会网络分析问题起源于物理学中的适应性网络，通过研究网络关系，有助于把个体间关系的“微观”网络与大规模的社会系统的“宏观”结构结合起来。从社会网络的角度出发，人在社会环境中的相互作用可以表达为基于关系的一种模式或规则，而基于这种关系的有规律模式反映了社会结构，这种结构的量化分析是社会网络分析的出发点。社会网络分析不仅仅是一种工具，更是一种关系论的思维方式。

9. 数据挖掘方法

数据挖掘(data mining)是从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的、人们事先不知道的但又是潜在有用的信息和知识的过程。这个定义本身就包括好几层含义：数据源必须是真实的、大量的、含噪声的；发现的是用户感兴趣的知

[12] 邱均平. 关于内容分析法的研究[J]. 中国图书馆学报，2004(2)：12—17.

[13] 蒲群莹. 国外内容分析项目研究[J]. 图书馆杂志，2005(4)：57—60.

[14] 范并思. 社会科学信息中的文本挖掘[J]. 图书情报工作，2012(4)：6—9.