情报学报 2017年3月 第36卷 第3期
Journal of the China Society for Scientific and Technical Information, March 2017, 36(3): 319-330

# 学术专著引用行为研究——基于引文内容特征分析的视角

章成志[1,2,3]，王玉琢[1,3]，卢 超[1,3]

（1. 南京理工大学经济管理学院信息管理系，南京 210094；2. 江苏省数据工程与知识服务重点实验室（南京大学），南京 210093；3. 江苏省社会公共安全科技协同创新中心，南京 210094）

**摘 要** 当前的引文内容分析研究基本上以学术论文为研究对象。与学术论文相比，学术专著的篇幅较长、引文内容更加翔实。学术专著的引文内容特征分布对学术专著的引文内容分析方法、乃至引用行为与动机等研究都具有重要的意义。本文首先通过对 Morgan & Claypool 出版的 39 本学术专著的引文内容进行人工标注，构建包含 13539 条引文内容的学术专著引文内容分析语料；然后分别从引文内容位置分布、引文提及次数以及引文内容上下文特征等三个方面分析学术专著的引文行为，并针对不同学科学术专著的引用行为特点进行了总结；最后本文讨论了学术专著与学术论文引文内容特征分布的差异，并指出进行学术专著引文内容分析时应注意的问题。本文研究成果可以帮助学者进一步了解学术专著引用行为的特点，并为后续的学术专著中的引文动机研究打下基础。

**关键词** 学术专著；引文内容分析；引用行为；引用位置；提及次数

# Citing Behavior of Academic Monographs—Perspective Based on Character Analysis of Citation Content

Zhang Chengzhi[1,2,3], Wang Yuzhuo[1,3] and Lu Chao[1,3]

(1. *Department of Information Management, Nanjing University of Science & Technology, Nanjing* 210094;
2. *Jiangsu Key Laboratory of Data Engineering and Knowledge Service (Nanjing University), Nanjing* 210093;
3. *Jiangsu Collaborative Innovation Center of Social Safety Science and Technology, Nanjing* 210094)

**Abstract:** Nowadays, the analysis of citation content is primarily based on academic papers. Comparing with academic papers, academic monographs have longer length and more informative citation content. So, the characteristics of citation content in academic monographs is of great importance to the study on analyzing method, even the behavior and motivation of citation content.This paper firstly mines citation content manually from 39 Academic monographs which published by Morgan & Claypool, and builds corpus that contains 13548 citation content of academic monographs. Secondly, the behavior of academic monographs is analyzed in three aspects: the location of citation content, mention of citation and characteristics of citation context, and then the features of citation behavior of monographs in different discipline are summarized. Lastly, this article also discusses differences between characteristic distribution of citation content in academic monographs and that in academic papers, as well as pointing out what should be paid more attention to when analyzing citation content in academic monographs. The result will be helpful to scholars' further knowing about c academic monograph's citation behavior, and lay a foundation of the follow-up research about citation motivation.

收稿日期：2016-08-10；修回日期：2017-01-08
基金项目：国家社科基金重大项目“面向知识创新服务的数据科学理论与方法研究”（16ZAD224）。
作者简介：章成志，男，1977年生，博士，教授，博士生导师，通讯作者，主要研究领域为信息组织、信息检索、数据挖掘及自然语言处理，E-mail: zhangcz@njust.edu.cn；王玉琢，女，1995年生，情报学专业硕士研究生，主要研究方向为信息检索与数据挖掘；卢超，男，1991年生，管理科学与工程专业博士生研究生，主要研究方向为引文内容分析。

**Key words:** academic monograph; citation content analysis; citation behavior; location of citation content; mention of citation

## 1 引 言

引文分析包含了基于著录信息的引文分析和基于引文内容的引文分析两大方面[1]。如今，随着科学技术的不断进步，全文数据库开始向用户免费开放，有了大规模的数据作为支撑，引文分析的范畴由简单的著录信息统计扩展到内容层面分析，基于引文内容的引用行为分析成为时下的研究热点。所谓引用行为，指的是“科研人员基本的学术行为，它可以反映研究人员借鉴之前研究成果的情况，展现科学的继承性和交流性，也是尊重他人科研成果的体现”[2]。

然而，目前的引文内容分析主要以学术论文作为研究对象，学术专著虽然同为学术交流的重要媒介，但关于其引用行为的研究并不多。由于学术专著在出版前往往已发表了相关的学术论文，加之学术专著多为纸质出版，电子专著的更新速度远远滞后于期刊数据库，此外，由于缺乏相应的学术专著全文数据库，人们可接触到的各类版本专著不利于数据处理，因此多数的学术专著引用研究还停留在学术专著被引情况分析上，仅通过被引频次的统计来揭示学术专著的学术研究价值。少量的学术专著施引分析也只停留在浅层的频次统计层面，用来验证学术专著在引文分析中的重要作用，关于一本学术专著的引文长度、位置、动机、作用等引用行为研究完全空白。以学术专著 *Apoptosis* 中的一段引文“Direct PS binding receptors are the T cell immunoglobulin and mucin (TIM) family (Kobayashi et al., 2007; Miyanishi et al., 2007; Santiago et al., 2007), brain angiogenesis inhibitor (BAI1) (Park et al., 2007a), and Stabilin-2 (Park et al., 2008).” ①为例，传统的专著引文分析只统计其中的参考文献出现频次以及来源和去向，这些不通过原文内容即可完成。而基于引文内容的专著引用行为分析，则需要通过上述引用的内容，一方面统计原句中的单词个数、句子长度、被引文频次、出现位置，另一方面通过其语义深究该引文句的作用、动机和上下文间的关系等。这些可对现有的引文内容研究领域进行扩充，帮助学者更好地了解专著引文的特点，从而促进引文推荐、引文检索以及各领域学术专著及作者的评价等工作的发展。

为了掌握学术专著引文内容分布的特点，为后续的学术专著引文内容分析方法提供依据，构建大规模的学术专著引文内容分析语料显得尤为必要。为此，本文人工标注了四个学科共计 39 本学术专著中的引文内容，构建包含了 13539 条引文内容的语料库，并利用该语料库分析了学术专著的引文内容位置、引文提及次数以及引文内容上下文特征，从而总结出学术专著的引用行为特点。

## 2 学术专著引用相关研究

学术专著在学术交流中扮演着十分重要的角色，一方面很多学者常常选择用学术专著的形式来展现自己的研究成果，其中所包含的引文信息对于研究知识交流等问题有着极为重要的意义。另一方面，学者们的在进行学术写作时，常常选择学术专著作为参考文献，研究学术专著的被引情况可以对学术专著的重要性进行衡量，并由此确定各领域中的核心专著。

### 2.1 学术专著引文分析

学者在进行学术专著撰写的过程中，常常要进行大量的引文引用工作，学术专著中的引文有着极为重要的研究意义。Bar-Ilan[3]通过对图书中的引文来源进行统计分析，发现了三大数据库(Google Scholar, Scopus, Web of Science)之间的数据互为补充的关系。Torres-Salinas 等[4]则根据 BKCI 中图书的引文内容，研究其中的引文来源和去向，从而提出了全新的引用模式——heliocentric clockwise maps。一些学者通过比较相同引文在学术论文和学术专著中所起的作用，来衡量学术专著在学术交流中的价值。Thompson[5]以 Choice 中“literature from the English and American Literature”模块 1995 年出版的图书为研究对象，利用 Arts and Humanities Citation Index 追踪其中的参考文献在学术论文中的被引情况，发现学术专著在洞悉学术出版模式中有重要意义。Kousha 等[6]以 IS&LS 学科分类中的 51 ISI-index 文献作为研究对象，观察其在 Google Book Search 图书

① Ning Yang, IngSwieGoping. Apoptosis[M]. American: Morgan & Claypool Life Sciences, 2013

和“Institute for Scientific Information Databases”论文中的被引情况，发现 Google Book Search 中的图书可以作为一项极佳的资源，以弥补 ISI 中图书引文不足的情况。随后 Kousha 等[7]又在前文研究的基础上提出了全新的 Google Book 引文自动抽取方法，并将所得结果与 BKCI 作对比，发现了二者之间极高的相关性，甚至在某些领域 Google Book 表现更优，从而进一步验证了图书引文的重要研究意义。

### 2.2　学术专著被引分析

早在 1972 年，Garfield[8]就提出可以利用引文被引分析作为工具，来衡量学术期刊的价值。与此相同，学术专著的影响力也可以通过其被引情况来衡量。Lindholm-Romantschuk 等[9]利用 Arts & Humanities Citation Index 和 the Social Sciences Citation Index，同时分析了各个学科中学术专著和学术论文的被引情况，并发现同一学科中的学术专著影响力比学术论文的影响力更高。苏新宁[10]则着眼于学术专著，统计了 CSSCI 中图书的被引频次，据此概括了我国人文社科类专著的被引情况，并给出了各学科最具影响力的 5 本专著。张希等[11]以 5 种图书馆学核心期刊为研究对象，分析统计其文献被引来源中的学术专著分布情况，根据学术专著被引频次了解专著在期刊论文写作中的影响力。Hammarfelt[12]则从高被引专著入手，分析了 Literature: General in Journal Info 中排名前两百的引文归属，从中挑选出被引用的学术专著，分析总结了其高被引原因。杨思洛等[13]将研究范围缩小到图书情报学领域，发现图书被引情况的年代分布符合该学科的整体发展情况。Torres-Salinas 等[14]利用 BKCI 中的 28634 本图书，分析并验证了影响图书被引率的三大因素，即：图书是否被编辑，图书是否成系列出版，以及图书出版商。将之与张希等[11]的研究结果结合可发现，高校工作者撰写的专著和高校出版社出版的专著，往往都会获得很高的被引率，对于学术交流的影响力很大。

总体来说，现有的学术专著引文分析，多是利用学术专著参考文献列表中的内容，来揭示与学术专著相关的数据库、出版商、检索系统之间的关系，以体现学术专著作为知识传播媒介在引文分析中的价值。学术专著被引分析也往往只利用引文索引或论文的参考文献列表进行简单的频次统计，概括出各学科领域专著被引情况，找出高被引学术专著与作者，并分析其高被引原因。目前尚无人研究学术专著的引文内容分布特征，而利用具体引文内容对学术专著的引用行为进行分析，则可知晓引文在学术专著中具体的长度、位置、动机、作用等信息，从而分析其特征，并将之运用到引文推荐、引文检索以及学术评价等具体实践中，具有重要的理论和实践意义。

## 3　学术专著引文内容数据的获取与标注

本文首先获取了 39 本学术专著的全文，设计了图书引文内容标注规则，根据此规则对 39 本学术专著进行引文内容的人工标注。

### 3.1　学术专著全文本获取

本文在 Morgan & Claypool 出版社①出版的图书中，从 Biomedical & Life Sciences 和 Engineering & Computer Sciences 两个类中各选两个子类：Building Blocks of the Cell-Cell Structure and Function; Integrated Systems Physiology—From Molecule to Function to Disease; Human-Centered Informatics; Visual Computing Computer Graphics, Animation, Computational Photography and Imaging。再从四个子类中随机选取 10 本（其中 Building Blocks of the Cell-Cell Structure and Function 类只包含 9 本专著）共计 39 本学术专著作为元数据，所选专著各章节内容前后连贯，并不独立成章，具体信息见表 1。

### 3.2　学术专著的引文内容标注

本文从章节信息与引文内容特征两个方面，对以上 39 本学术专著的引文内容进行人工标注，具体的标注规范及字段如表 2 所示，图 1 给出了图书引文内标注样例。

对 39 本学术专著进行标注后，共得到 13539 条引文正文，8031 条引文前第二句，10529 条引文前第一句，10138 条引文后第一句，7829 条引文后第二句。

### 3.3　学术专著章节信息统计

首先以学术专著为研究单位，统计每本学术专著的总章数，结果如图 2 与表 3 所示。

① http://www.morganclaypool.com/

**表1 实验数据的元数据说明表**

| 类别 | Biomedical & Life Sciences | | Engineering & Computer Sciences | |
|---|---|---|---|---|
| 子类 | Building Blocks of the Cell-Cell Structure and Function(简称 S1) | Integrated Systems Physiol-ogy—From Molecule to Func-tion to Disease(简称 S2) | Human-Centered Informatics (简称 S3) | Visual Computing Computer Graphics, Animation, Computational Photography and Imaging(简称 S4) |
| 学术专著编号及书名称 | 1 Apoptosis | 10 Alveolar Structure and Function | 20 Adaptive Interaction A Utility Maximization Approach to Understanding Human Interaction with Technology | 30 Gazing at Games An Introduction to Eye Tracking Control |
| | 2 $Ca^{2+}$-dependent Signal Transduction | 11 Capillary Fluid Exchange Regulation, Functions, and Pathology | 21 Constructing Knowledge Art An Experiential Perspective on Crafting Participatory Representations | 31 High Dynamic Range Video |
| | 3 Epithelial Polarity | 12 Cardiovascular Responses to Exercise | 22 Core-Task Design A Practice-Theory Approach to Human Factors | 32 Information Theory Tools for Computer Graphics |
| | 4 Intermediate Filaments | 13 Developmental Programming of Cardiovascular Disease | 23 Designing and Evaluating Usable Technology in Industrial Research Three Case Studies | 33 Information Theory Tools for Image Processing |
| | 5 Membrane Nanodo-mains | 14 Enteric Nervous System—The Brain-in-the-Gut | 24 Geographical Design Spatial Cognition and Geographical Information Science | 34 Interactive Shape Design |
| | 6 Phagocytosis | 15 Erectile Dysfunction | 25 How We Cope with Digital Technology | 35 Mathematical Tools for Shape Analysis and Description |
| | 7 Protein Translation | 16 Hemorheology and Hemodynamics | 26 Interacting with Information | 36 Real-Time Massive Model Rendering |
| | 8 The Actin Cy-toskeleton and the Regulation of Cell Migration | 17 Intestinal Immune System | 27 Making Claims Knowledge Design, Capture, and Sharing in HCI | 37 Virtual Crowds Methods, Simulation, and Control |
| | 9 Vesicular Transport in the Secretory and Endocytic Pathways | 18 Peripheral Arterial Dis ease—Pathophysiology and Therapeutics | 28 Proxemic Interactions From Theory to Practice | 38 Virtual Crowds Steps Toward Behavioral Realism |
| | | 19 Regulation of Vascular Smooth Muscle Function | 29 Surface Computing and Collaborative Analysis Work | 39 Wang Tiles in Computer Graphics |

**表2 学术专著引文标注信息表**

| 标注类别与标识 | | 具体含义 |
|---|---|---|
| 类别 | 标识 | |
| 章节信息 | Chapter_No | 引文所在章号 |
| | Chapter | 引文所在章名称 |
| | Section_No | 引文所在节号 |
| | Section | 引文所在节名称 |
| | Chapter_Len | 章节长度，即引文所在章中所有单词数 |
| 引文内容特征 | Citation_No | 引文内容编号，按章节计算，每章的第一句引文为1 |
| | Citation_Location | 引文内容位置，引文第一个单词所在的位置，按章统计 |
| | Citation_Content | 引文具体内容，不包括参考文献标注 |
| | Reference | 引文内容对应的参考文献 |
| | Previous_2 | 引文往前数第二句 |
| | Previous_1 | 引文前一句 |
| | After_1 | 引文后一句 |
| | After_2 | 引文往后数第二句 |

计算39本学术专著的最大章数、最小章数、中间章数以及平均章数，发现章节数最大的为24章，最小的为4章，超过15章的专著有2本，低于5章的专著有3本，出现次数最多的章节数是6和7，出现次数最少的章节数是16和24，即大多数学术专著的章节数保持在5~15章（包括5章和15章），极高和极低的情况都是个例。

在前文研究的基础上，对两大类共计四个学科方向的专著分别进行统计，统计每学科专著的平均章数，结果如表4所示。由表4可知，各个类别之间的平均章节数并无太大差别，Integrated Systems Physiology—From Molecule to Function to Disease 和 Visual Computing Computer Graphics, Animation, Computational Photography and Imaging 两个类目的平均章节数偏高主要是因为章节数最多的两本书（16章和24章）分别出现在其中，总体来说，学术专著的章节数并不因为所处的学科不同而有所差别。

上述结果表明，虽然学术专著和学术论文都通过不同的章节安排来进行学术成果展示。但学术专著的章节数长短不一，并无固定的标准，各章节对应的具体内容也不固定。而相比之下，学术论文的章节安排则更为固定，最常见的格式为四节 IMRaD (introduction, methods, results, discussion)[15]，各章节的具体内容较为固定，但也会出现交叉。

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Chapter_No | Chapter | Sectio | Section | Chapter_ | Cita | Citat | citation_content | reference | Previous_2 | Previous_1 | After_1 | After_2 |
| 2 | chapter1 | Phagocyt | 1.1 | INTRODUCTIO | 970 | 1 | 211 | In addition, the | Aderem, A., and | | | While the term | On the other |
| 3 | chapter1 | Phagocyt | 1.2 | FUNCTIONAL | 970 | 2 | 250 | Phagocytosis is | Silverstein, | | | Epithelial cells | However, these |
| 4 | chapter1 | Phagocyt | 1.2 | FUNCTIONAL | 970 | 3 | 335 | In addition, the | Rabinovitch, M. | Epithelial cells | However, these | On the other | These " |
| 5 | chapter1 | Phagocyt | 1.2 | FUNCTIONAL | 970 | 4 | 402 | Upon interaction | Rabinovitch, M. | On the other | These " | Following | Studies have |
| 6 | chapter1 | Phagocyt | 1.2 | FUNCTIONAL | 970 | 5 | 425 | Following | Pfeifer, J.D., | These " | Upon interaction | Studies have | Large |
| 7 | chapter1 | Phagocyt | 1.2 | FUNCTIONAL | 970 | 6 | 470 | Large | Savill, J. 1997. | Following | Studies have | Phagocytosis is | |
| 8 | chapter1 | Phagocyt | 1.2 | FUNCTIONAL | 970 | 7 | 487 | Phagocytosis is | Gaipl, U.S., S. | Studies have | Large | | |
| 9 | chapter1 | Phagocyt | 1.2 | FUNCTIONAL | 970 | 8 | 563 | Production of | Griffin, F.M. | A comprehensive | What | | |
| 10 | chapter1 | Phagocyt | 1.2 | FUNCTIONAL | 970 | 9 | 563 | Production of | Platt, N., R.P. | A comprehensive | What | | |
| 11 | chapter1 | Phagocyt | 1.3 | OVERVIEW OF | 970 | 10 | 639 | Macrophages, | Griffin, F.M., | | The specificity | A class of | In contrast, |
| 12 | chapter1 | Phagocyt | 1.3 | FUNCTIONAL | 970 | 11 | 671 | A class of | Janeway, C.A. | The specificity | Macrophages, | In contrast, | Opsonins coat |
| 13 | chapter1 | Phagocyt | 1.3 | OVERVIEW OF | 970 | 12 | 756 | Among the well- | Anderson, C.L., | In contrast, | Opsonins coat | The other major | |
| 14 | chapter1 | Phagocyt | 1.3 | FUNCTIONAL | 970 | 13 | 789 | The other major | Ross, G.D., W. | Opsonins coat | Among the well- | | |
| 15 | chapter1 | Phagocyt | 1.4 | PARTICLE | 970 | 14 | 912 | F-actin | Kaplan, G. 1977. | In contrast, | This latter | Following | |
| 16 | chapter1 | Phagocyt | 1.4 | PARTICLE | 970 | 15 | 933 | Following | Desjardins, M., | This latter | F-actin | | |
| 17 | chapter2 | Innate | | | 1044 | 1 | 35 | In comparison to | Aderem, A., and | | The phagocytic | Additionally, | As a simple |
| 18 | chapter2 | Innate | | | 1044 | 2 | 233 | Therefore, the | Janeway, C.A. | Additionally, | As a simple | | |
| 19 | chapter2 | Innate | 2.1 | PATT ERN | 1044 | 3 | 291 | This makes it | Janney, P.A., | | Macrophages are | The well-studied | While some of |
| 20 | chapter2 | Innate | 2.1 | PATT ERN | 1044 | 4 | 351 | While some of | Aderem, A., and | This makes it | The well-studied | Carbohydrates on | Mannose |
| 21 | chapter2 | Innate | 2.1 | PATT ERN | 1044 | 5 | 386 | Carbohydrates on | Ezekowitz, R.A., | The well-studied | While some of | Mannose | Another lectin |
| 22 | chapter2 | Innate | 2.1 | PATT ERN | 1044 | 6 | 386 | Carbohydrates on | Peiser, L., P.J. | The well-studied | While some of | Mannose | Another lectin |
| 23 | chapter2 | Innate | 2.1 | PATT ERN | 1044 | 7 | 421 | Mannose | Stahl, P.D., and | While some of | Carbohydrates on | Another lectin | TLRs bind to a |
| 24 | chapter2 | Innate | 2.1 | PATT ERN | 1044 | 8 | 470 | Another lectin | Brown, G.D., and | Carbohydrates on | Mannose | TLRs bind to a | TLRs are also |
| 25 | chapter2 | Innate | 2.1 | PATT ERN | 1044 | 9 | 496 | TLRs bind to a | Underhill, D.M., | Mannose | Another lectin | TLRs are also | Another class of |
| 26 | chapter2 | Innate | 2.1 | PATT ERN | 1044 | 10 | 523 | TLRs are also | O' Neill, L.A.J. | Another lectin | TLRs bind to a | Another class of | Pattern |
| 27 | chapter2 | Innate | 2.1 | PATT ERN | 1044 | 11 | 523 | TLRs are also | Martinon, F., K. | Another lectin | TLRs bind to a | Another class of | Pattern |
| 28 | chapter2 | Innate | 2.1 | PATT ERN | 1044 | 12 | 556 | Another class of | Shaw, M.H., T. | TLRs bind to a | TLRs are also | Pattern | Scavenger |

图 1　图书引文标注结果样例

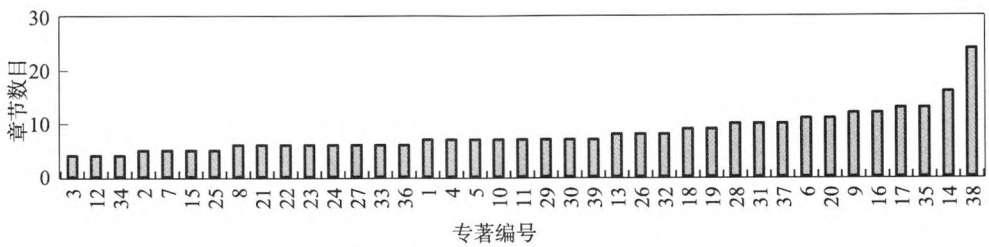


图 2　学术专著章节数分布图

**表 3　学术专著章节数相关指数表**

| 类别 | 最大章节数 | 最小章节数 | 中间章节数 | 平均章节数 |
|---|---|---|---|---|
| 数量/章 | 24 | 4 | 7 | 8.21 |

**表 4　学科章节数统计表**

| 类别 | 专著编号 | 章数 | 专著编号 | 章数 | 平均数 |
|---|---|---|---|---|---|
| S1 | 1 | 7 | 2 | 5 | 7.11 |
| | 3 | 4 | 4 | 7 | |
| | 5 | 7 | 6 | 11 | |
| | 7 | 5 | 8 | 6 | |
| | 9 | 12 | | | |
| S2 | 10 | 7 | 11 | 7 | 9.00 |
| | 12 | 4 | 13 | 8 | |
| | 14 | 16 | 15 | 5 | |
| | 16 | 12 | 17 | 13 | |
| | 18 | 9 | 19 | 9 | |
| S3 | 20 | 11 | 21 | 6 | 7.10 |
| | 22 | 6 | 23 | 6 | |
| | 24 | 6 | 25 | 5 | |
| | 26 | 8 | 27 | 6 | |
| | 28 | 10 | 29 | 7 | |
| S4 | 30 | 7 | 31 | 10 | 9.50 |
| | 32 | 8 | 33 | 6 | |
| | 34 | 4 | 35 | 13 | |
| | 36 | 6 | 37 | 10 | |
| | 38 | 24 | 39 | 7 | |

注：学术专著编号见表 1。

## 4　学术专著的引文行为分析

本文根据 39 本学术专著的人工标注结果，分别从引文内容位置分布、引文提及次数以及引文内容上下文特征等三个方面进行学术专著的引用行为分析。

### 4.1　引文内容位置分布

#### 4.1.1　引文在全文的位置分布

本文通过 citation-location 和 chapter-len 计算出每条引文在全文中的绝对位置以及每本专著的绝对长度，两值相除得出每条引文在专著全文中的相对位置，39 本学术专著的引文全文位置分布情况如下图。由图 3 中散点分布情况可见，引文在学术专著全文中的分布较为均匀，并未出现明显的集中分布点。

#### 4.1.2　引文在各章节的位置分布

（1）绝对位置分布

首先，根据所记录数据中引文的绝对位置，以专著作为单位，计算每本专著每一章中引文的平均绝对位置值，所得结果分别如图 4 所示。由图 4 可见，学术专著中大部分引文的绝对位置在 4000 以下，多数集中在 0~2000，6000 以上最少。引文绝对位置在 0~2000 最集中的是最后一章，在 2000~4000 最集中的是 0.5 章，即学术专著的最中间章节，在 4000

以上的引文，则多数分布在0.5章节以后，即专著的后半部分。总体来说，学术专著各章节的引文数量与绝对位置之间呈反比关系，绝对位置越大包含的引文数越少，而高绝对位置的引文，则主要出现在专著的后半部分。导致该结果的可能原因是，随着专著内容的推进，章节的篇幅越来越大，因此后半部分会出现绝对位置更高的引文。

随后，在前文结果的基础之上，对于每一学科类专著，计算所有相同章节中引文出现的平均绝对位置，结果如图5所示。由图5结果可以看出，不同学科各章节的引文平均绝对位置存在高峰期，但每类学科的高峰期又不尽相同，四个学科的分布折线找不到至少包含三个学科的重合点，整个引文平均绝对位置的分布并未出现过多的重叠，这说明无论是同一大类不同学科的专著，还是不同大类的专著，彼此之间引文的绝对位置分布都是相互独立的，并无明显的共有规律。

（2）相对位置分布

将引文在章节内部的相对位置定义为预处理数据表citation-location/chapter-len的值。

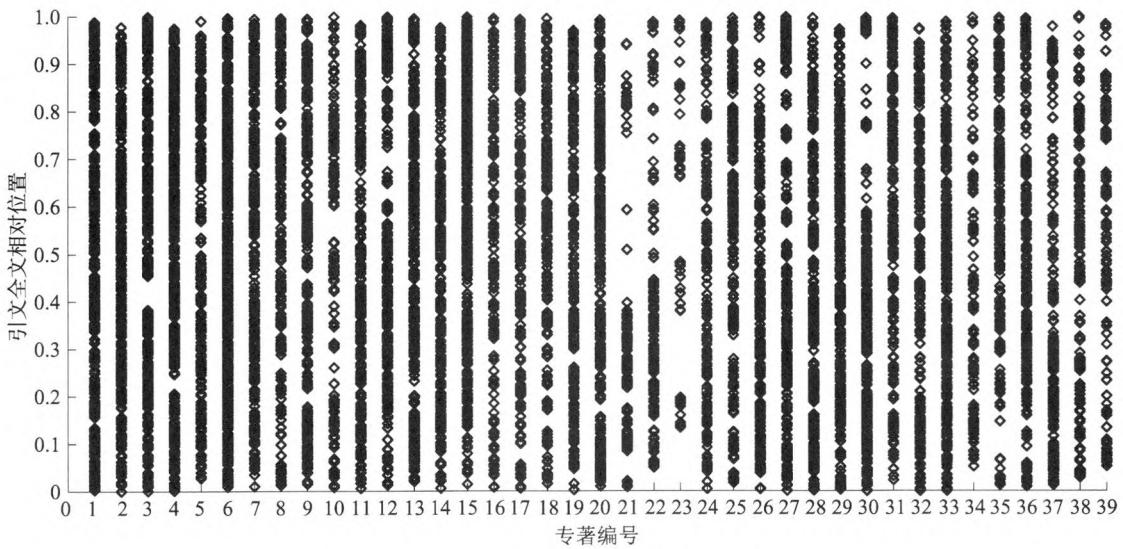


图3 学术专著全文引文相对位置散点图

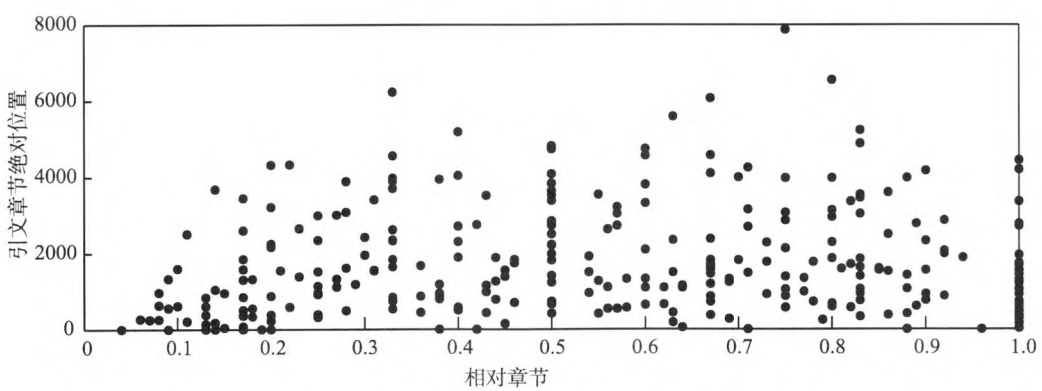


图4 学术专著各章节引文平均绝对位置散点图

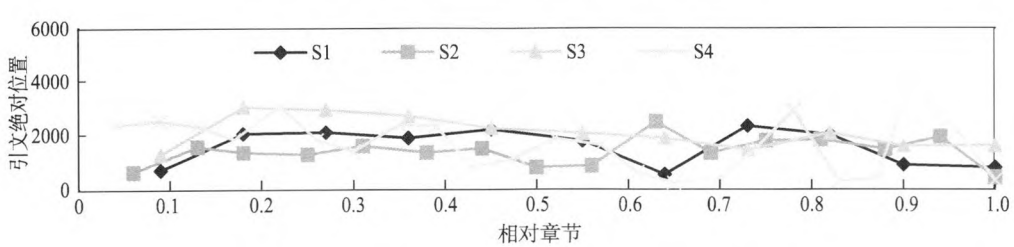


图5 不同学科各章节引文平均绝对位置分布图

图 6 反映了学术专著章节内部的引文分布情况，每一列为每本专著每一章节内部的引文分布情况。由图中散点分布情况可见，引文在专著各章节的分布比较均匀，并未出现十分集中的分布区域，这与引文在专著全文中的分布情况相呼应，即专著中的引文分布，无论是在以全书还是以章节为单位，都比较均匀，无明显的集中分布区。

不同学科中相同章节的平均引文相对位置分布情况如图 7 所示。由图 7 可知，同属一大类的 S1 与 S2 在 0.1~0.5 章，即学术专著的前半部分的引文相对位置极为接近，而 S3 与 S4 在前 0.5 章内容中的引文相对位置较之后 0.5 章也更为接近。而不同类别的学科之间则未出现明显的共同趋势。总体而言，相同类别的不同学科专著，在全书的前半部分内容中引文分布位置较为接近，而不同类别的学科之间引文章节内部的位置分布并无关联。

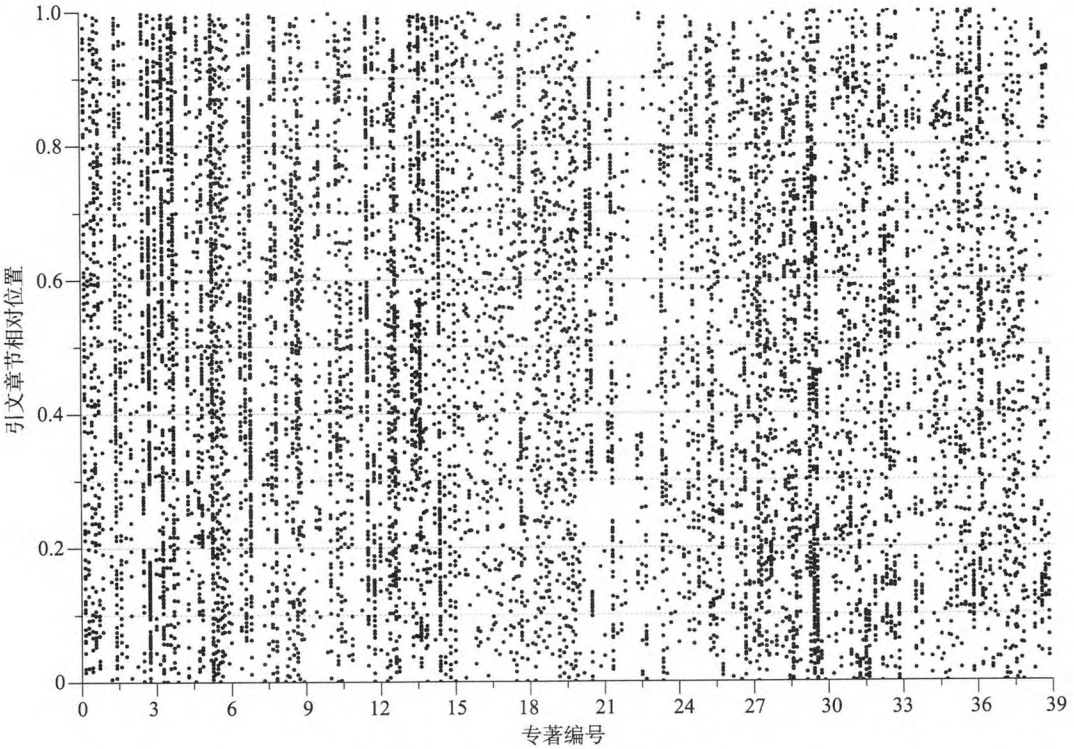


图 6　学术专著各章节引文相对位置散点图

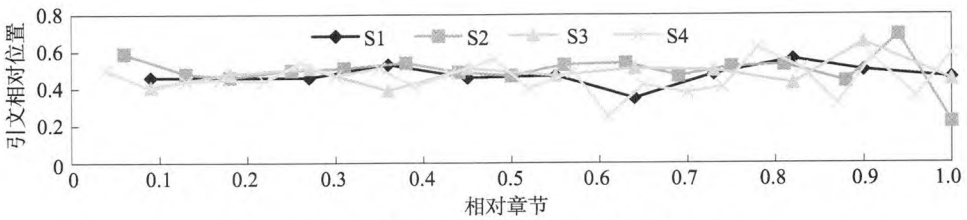


图 7　不同学科各章节引文平均相对位置分布图

## 4.2　引文提及次数

（1）引文在全文的提及次数

针对每本专著，从全文的角度出发，通过查找其原著末尾的参考文献列表和预处理数据表中统计所得的引文总数，计算每本专著的引文提及总数与参考文献数的比值，由此分析专著中是否存在文献多被引情况。将每本专著的引文提及总次数、参考文献总数和二者比值生成统计图，分别如图 8、图 9 所示。可见 39 本专著的引文提及次数和参考文献数量的分布并不是完全的正比关系，随着参考文献数的增加，引文提及次数分布出现了多处波折。但每本专著的引文提及次数和参考文献数的比值都超过了 1，多数专著这一比值集中在 1~1.5，9 本专著的则超过了 2，由此可见参与统计的每本学术专著中都出现了引文重复被引现象。

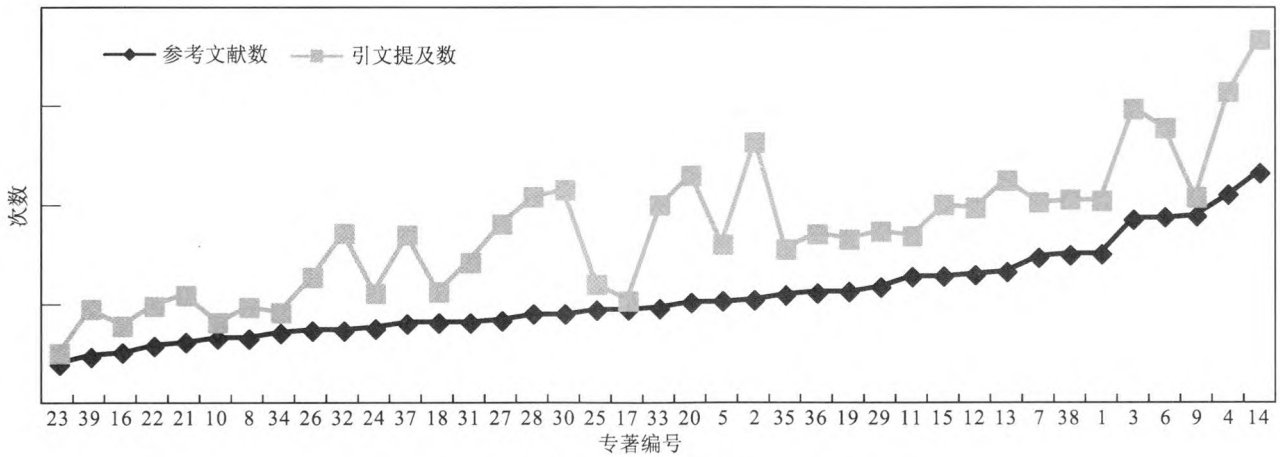


图 8 专著的引文提及次数和参考文献总数统计图

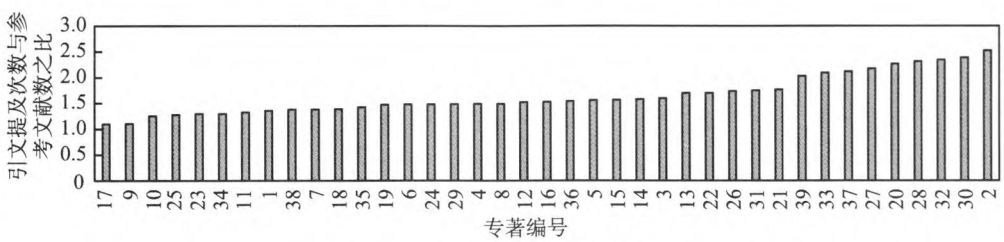


图 9 各专著引文在全文中提及情况分布图

同时，对于两大类别的四个学科，分别计算其中所有专著引文提及次数与参考文献数比值的平均值，表 5 中结果表明，每一学科专著的参考文献都存在重复被引的情况，同属一大类的两个学科 S1、S2 的引文提及次数与参考文献数比值仅相差 0.02，同样 S3、S4 之间的这一比值也仅相差 0.02。而将两大类作对比，则 Engineering & Computer Sciences 类的比值明显高于 Biomedical & Life Sciences 类。综上所述，相同类别的学术专著的引文提及情况十分接近，而不同类别专著的引文提及情况则有着明显的区别。

**表 5 各学科引文在全文中提及情况统计表**

| 学科 | S1 | S2 | S3 | S4 |
|---|---|---|---|---|
| 平均引文提及次数 | 450.67 | 340.9 | 282.7 | 324.6 |
| 平均参考文献数 | 313.11 | 240.9 | 161.4 | 183.1 |
| 比值 | 1.44 | 1.42 | 1.75 | 1.77 |

（2）引文在各章节的提及次数

本文在预处理结果数据表中，利用数据透视表，先分别统计每章节中的每篇参考文献提及次数，随后求平均值。39 本学术专著的每章节平均参考文献提及次数结果如图 10、图 11 所示。图 10 表明，每本专著各章节的平均提及次数在 1.00~1.50 最多，1.50~2.00 次之，这与引文提及次数在专著全文中的分布情况相同。平均提及次数超过 2.00 的引文

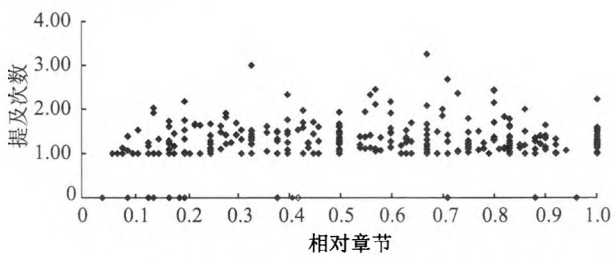


图 10 学术专著各章节平均引文提及次数散点图

则主要分布在 0.5 以后的章节。由图 11 可见，39 本专著里，每本专著最高提及次数出现最多的章节是相对第 0.3 章，最低提及次数出现最多的章节是相对第 1 章，即最后一章。以纵坐标 50%作为分界线，出现在分界线上最多的区间是相对章节 0.5~0.6。上述结果表明，重复被引次数较高的文献主要出现在学术专著的后半部分内容中，并且多数学术专著全书后半部分内容中的平均引文提及次数比前半部分大。

针对不同学科，统计其中所有专著相同章节的参考文献提及次数的平均值，其中同属一大类的 S1、S2 总体情况较为接近，各章节平均提及次数除去个别偏高外，主要分布在 1.00~1.50。而不同类别的学科，S1、S2、S3 在前 0.5 章中，S1、S2、S4 则在后 0.5 章中的平均提及次数都较为接近。则对于相同类别的学科而言，存在某些类别其中专著的各章引文提及次数都很接近，而不同类别的不同学科专著也

在专著的不同部分，存在引文提及次数类似的情况。具体结果如图 12 所示。

### 4.3　引文内容上下文特征

本文针对预处理结果数据表中的引文及其前一、二句和后一、二句的内容，利用函数分别统计每本专著中上述五项内容的单词数和字符串长度，随后按章节求取平均值。所得结果如图 13、图 14 所示。图 13 表明，每本专著各章节的引文及其上下文的平均单词个数主要分布在 40.00 以下，其中 20.00~30.00 之间最多，平均单词数超过 40.00 的引文及其上下文则主要分布在 0.5 以前的章节中。由图 14 可见，39 本专著各章节的引文及其上下文的平均长度主要分布在 100.00~200.00，平均长度超过 200.00 的引文及其上下文同样主要分布在 0.5 以前的章节中。该结果表明，无论是引文正文还是引文上下文，其内容的单词数和长度都存在着共同的集中分布区，不因所属身份的变化而变化。并且高单词数和高长度的句子都分布在全书的前半部分中，且多数都为引文前第一句和引文后第二句。

随后在前文结果的基础上，针对不同学科，统计其中所有专著的引文及其上下文的平均单词数和长度，所得结果统计在表 6、表 7 中。由表 6 可见，每一学科中，除去 S3 中引文内容单词数偏大，S4 中的引文后第二句平均单词数偏小外，引文及其上下文之间的平均单词数差别并不大。而不同学科之间相

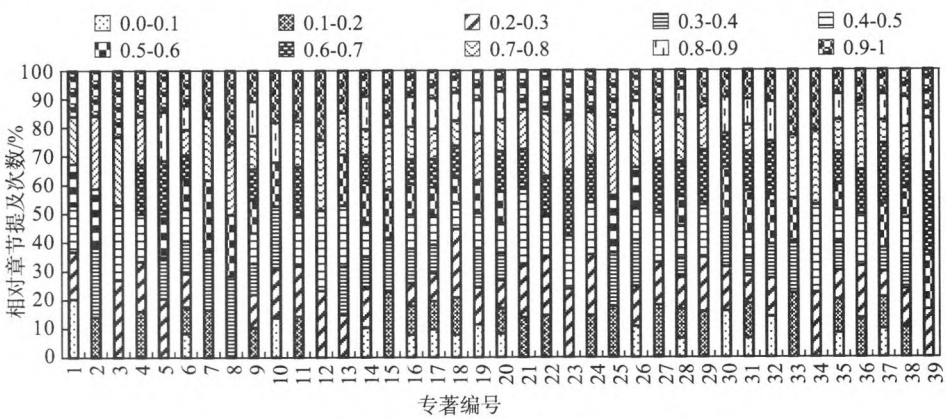


图 11　学术专著各章节平均引文提及次数柱状图

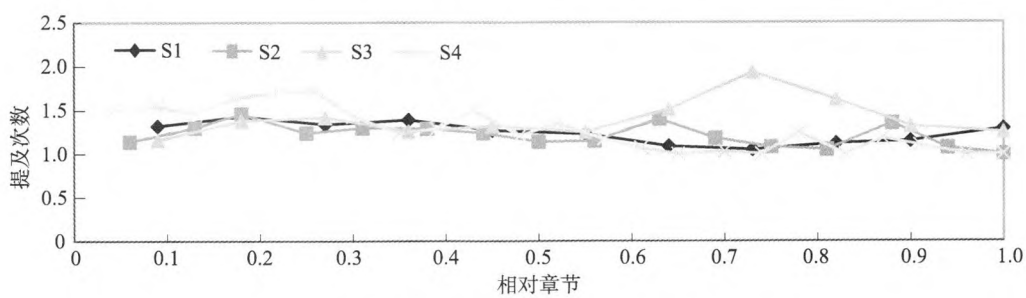


图 12　不同学科各章节平均引文提及次数分布图

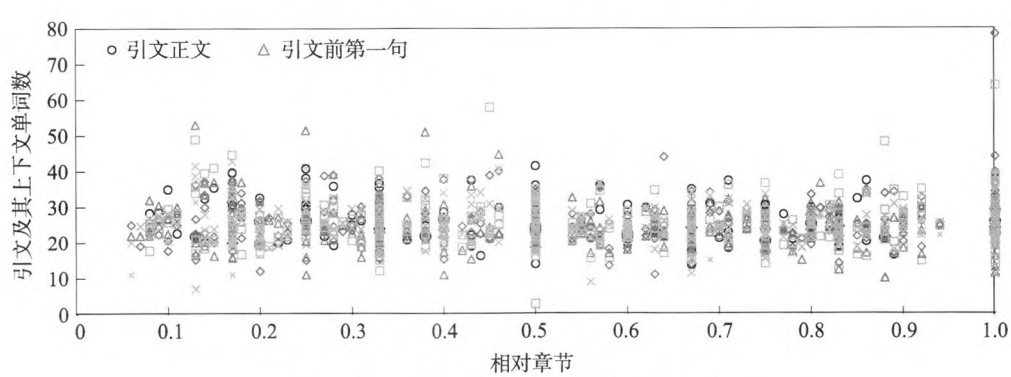


图 13　引文及其上下文平均单词数分布散点图

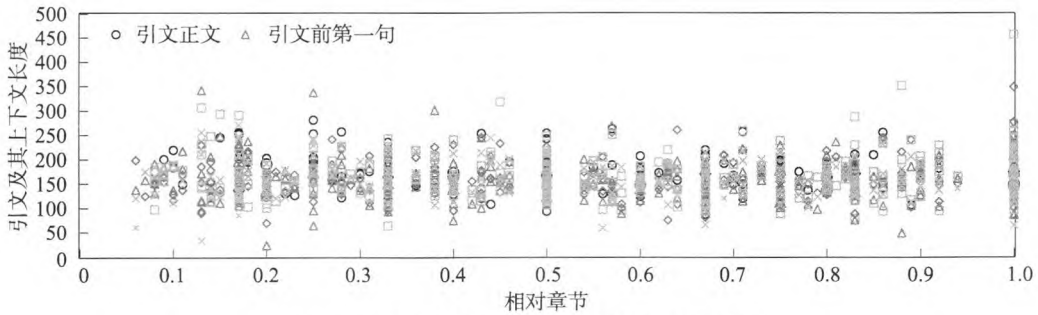


图 14 引文及其上下文平均长度分布散点图

同类目的平均单词数差距也仅在 5 个左右浮动。表 7 中各个学科，除去 S3 的差别较大，其他学科内引文及其上下文的平均长度差别在 15 以内浮动，较为接近。同属一大类的两学科间，S1 的所有数值皆高于 S2，S3 的各项数值皆高于 S4，S4 的引文及其上下文的平均长度都低于其他三个学科。总体来说，不同学科专著的引文及其上下文平均单词数差别并不大，而平均引文长度则存在着一定的差别。

**表 6 各学科学术专著引文及其上下文平均单词数分布表**

| 分类 / 学科 | 引文内容 | 引文前第二句 | 引文前第一句 | 引文后第一句 | 引文后第二句 |
|---|---|---|---|---|---|
| S1 | 29.42 | 26.53 | 26.92 | 27.08 | 27.72 |
| S2 | 25.27 | 24.35 | 24.99 | 24.49 | 23.58 |
| S3 | 30.82 | 25.16 | 26.79 | 26.94 | 27.19 |
| S4 | 26.47 | 26.50 | 29.33 | 29.00 | 22.40 |

**表 7 各学科学术专著引文及其上下文平均长度分布表**

| 分类 / 学科 | 引文内容 | 引文前第二句 | 引文前第一句 | 引文后第一句 | 引文后第二句 |
|---|---|---|---|---|---|
| S1 | 177.26 | 176.19 | 182.60 | 180.47 | 187.18 |
| S2 | 172.70 | 160.61 | 162.82 | 160.80 | 158.89 |
| S3 | 200.01 | 161.43 | 166.69 | 170.95 | 175.30 |
| S4 | 154.22 | 140.03 | 147.44 | 150.11 | 154.71 |

## 5 讨 论

同作为学术交流的重要媒介，学术专著和学术论文的引用行为都有着极为重要的研究意义，二者之间有很多相似点，但也存在着不同之处，本文将从以下几点进行比较，并针对其中的显著异同点讨论其产生原因。

### 5.1 学术专著与学术论文引文内容特征分布的定性比较分析

（1）引文位置分布

就绝对位置而言，学术专著各章节中的引文绝对位置主要集中在 0~2000，绝对位置越大包含的引文数量越少，从专著全文的角度来看，引文在全文和章节中的分布都较为均匀，未出现明显的引文集中分布点。不同大类的学术专著，其引文的相对位置和绝对位置分布皆无明显的共同之处。相比之下，学术论文的引文位置分布从全文角度出发，引文多在引言和背景综述部分最为密集，多数引文分布在全文的前 30%的内容中，而各章节内部的引文分布则比较均匀，并无明显的集中点[16]。

（2）引文提及次数

根据前文数据分析结果可知，39 本学术专著的引用个数全部高于引文篇数，但二者之间并未出现明显的比例关系，全文和各章节的引文平均提及次数多数都集中在 1.00 ~ 1.50 之间，高重复被引则更多的出现在一本专著的后半部分，且多数学术专著后半部分内容中的平均引文提及次数比前半部分大。相比之下，胡志刚等[17-18]的研究结果则表明，学术论文中引用次数和引文篇数之间则有较强的正相关性，其中引文平均提及次数在 1.5 次左右，而在不同章节中，引文的平均提及次数并没有出现太大的差异。

（3）引文及其上下文长度

学术专著中各章节的引文及其上下文的平均单词个数主要为 20.00~30.00 个，平均长度主要分布在 100.00~200.00，高单词数和高长度的句子都主要分布在全书的前半部分。不同学科专著的引文及其上下文平均单词数差别不大，而句子长度则存在着一定的差别。Singh 等[19]对 2600 多万条引文内容进行统计，所得结果表明学术论文中的引文平均单词数为 26 个，这正好处于学术专著 20~30 的范围之中。卢超等[20]则在文章中指出，学术论文的平均引文长度集中在 50~200。由此可见，学术专著和学术论文的平均引文单词数以及平均引文长度比较接近。

综上所述，学术专著和学术论文的引用行为相同点为：①引文的平均提及次数主要都在 1~1.5。②引文内容的单词数主要集中在 20~30 个，长度集中在 100~200。二者引用行为的不同点包括：①学术专

著的引文多出现在全书的前 60%内容中，学术论文中的引文则集中在全文发前 30%内容中。②学术专著的引文提及次数与参考文献篇数不成比例关系，而学术论文中则成正比关系。③学术专著各章节内部的引文位置集中在五分之三处、提及次数出现差异，而学术论文各章节之间的引文位置和提及次数并无差别。

针对上述的异同点，本文从以下三个方面分析原因。

（1）载体的篇幅差异

学术专著和学术论文的篇幅长度有很大的不同，由于篇幅限制，学者在撰写学术论文时，章节数太少则内容无法系统阐述，章节数太多则显得拖沓，因此学术论文的章节数较少且固定。而学术专著的篇幅一般都在百页以上，在结构安排上限制较少，学者们可以根据内容自行安排，这就导致学术专著章节的分布随机性很大。

（2）二者的功能异同

学术专著和学术论文的重要功能都是对学者的研究成果进行展示和传播,因此二者引文的单词数和长度极为接近，当中的引文内容是学者观点和思想的有力佐证，将之集中安排在全文的开篇部分能够尽早向读者证明自身研究的可靠性。但细究二者的功能可以发现，“学术论文的功能侧重于研究成果公示，而学术专著则侧重于成果传播”[21]。公示需要读者尽快接受其结果，引用内容作为佐证需要更早地出现在文章中，因此学术论文中的引文集中出现位置比学术专著更加靠前。

（3）作者的行文习惯

不同的作者的不同写作习惯导致了其不同的引文方式。学者在撰写学术专著和学术论文的过程中，可能因为引用习惯的不同，导致学术论文的引文提及次数和参考文献之间成正比关系，而学术论文中则相反。同时，相比于小规模的论文撰写，学者在撰写学术专著时，自由发挥的空间大，不同领域学者的不同写作习惯就会对行文产生较大的影响，这也就导致了不同学科专著的引用行为，一部分存在规律可循，一部分则表现得相对随机。

### 5.2　学术专著引文内容分析应注意的问题

前文结果表明，学术专著和学术论文的引用行为存在明显的异同点，学术论文的引文内容分析主要包括 Why、What、Where、Who、When、How 六大研究范畴，学术专著亦可从这六点入手，来分析其引文的引用动机、引文主题、引文位置、引用主体、引文发展规律和引用强度。二者的引文内容单词数和长度极为接近，因此在利用具体引文内容进行引用动机、引文主题等方面的研究时，学术专著可参照学术论文运用自然语言处理和文本挖掘技术来探讨其引用动机，同时利用词频统计、主题分类、主题可视化等方法来揭示引文主题和知识传播的特点。

除了上述共同之处外，由于学术专著中的全文引文集中分布位置比学术论文更广，各章节内部的引文又集中分布于中间位置，因此在研究学术专著的引文内容时，因将重点范围扩充得更广，并且更多地放在各章节的中间位置。同时，由于学术论文篇幅较短，其所研究的范围往往是全文，而学术专著的篇幅较长，各个章节的长度往往和一篇学术论文相当，因此除了全文范围的研究外，还应考虑单独章节内部的引文内容分析，对不同章节的引用行为特征进行横向比较，探讨其中的规律。此外，由于学术论文的章节安排较为固定，而学术专著的章节数量长短不一，因此在进行引用行为分析时，应注意将章节做归一化处理，从而在将不同的专著进行纵向比较时能使得比较的范畴相一致。

## 6　结　论

随着科技的不断进步，全文数据库的不断开放，基于全文内容的引用行为分析成为当下的研究热点,本文抛开传统的研究模式,以 Morgan & Claypool 出版社出版的 39 本学术专著为研究对象，通过对于其中引文内容的挖取，对学术专著的部分引用行为进行探究，发现学术专著中引文均匀分布在全文内容中，每章节中的引文也同样多为均匀分布。而每一本学术专著中都存在引文重复被引情况，但引文提及次数和参考文献篇数之间并无明显比例关系，引文重被引次数高的章节则常常出现在整本书的后半部分，且后半部分平均提及次数高于前半部分。引文及其上下文的平均单词个数主要分布在 20.00~30.00，平均长度主要分布在 100.00~200.00，高单词数和长度的引文及其上下文则主要分布在一本专著的前半部分内容中。上述行为特点与学术论文引用行为的异同原因可总结为载体的篇幅差异、学术专著与学术论文的功能区别以及作者的行文习惯影响。

本文只是对学术专著引用行为的初涉性研究，虽然得到了一些结果，但仍然存在很多不足之处。一方面参与研究的学术专著仅涉自然科学，基础数据的匮乏导致了在分析不同学科专著的引用行为

时，无法得出更为确切的规律。另一方面，对于引文位置的分布和引文提及次数的考察，仅考虑了单篇被引文献，得出的结果较为片面。此外，本文的研究虽然以引文内容为基础，但研究内容仅涉及句法层面，仅进行了一些简单的频次统计分析工作。后期研究可以以此为基础，考虑扩大所研究学科的范围，更为深入地分析高被引文献、共被引文献等在学术专著中的分布情况，并将句法研究延伸到语义层面上，借助文本挖掘、自然语言处理等方法，分析学术专著引文主题、作者的引用动机、引文的引用作用等，此外，后续还会将学术专著与学术论文的引文内容进行量化比较分析，对学术专著的引用行为进行更加广泛、深入的研究。

## 参 考 文 献


[1] 刘盛博, 丁堃, 张春博. 引文分析的新阶段: 从引文著录分析到引用内容分析[J]. 图书情报知识, 2015(3): 25-34.

[2] 邱均平. 科研人员论文引用动机及相互影响关系研究[J]. 图书情报工作, 2015, 59(9): 36-44.

[3] Bar-Ilan J. Citations to the ''Introduction to informetrics'' indexed by WOS, Scopus and Google Scholar[J]. Scientometrics, 2010, 82(3): 495-506.

[4] Torres-Salinas D, Rodríguez-Sánchez R, Robinson-García N, et al. Mapping citation patterns of book chapters in the Book Citation Index[J]. Journal of Informetrics, 2013, 7(2): 412-424.

[5] Thompson J W. The death of the scholarly monograph in the humanities? Citation patterns in literary scholarship[J]. International Journal of Libraries and Information Studies, 2002, 52: 122-136.

[6] Kousha K, Thelwall M. Google book search: Citation analysis for social science and the humanities[J]. Journal of the Association for Information Science and Technology, 2009, 60(8): 1537- 1549.

[7] Kousha K, Thelwall M. An automatic method for extracting citations from Google Books[J]. Journal of the Association for Information Science and Technology, 2015, 66(2): 309-320.

[8] Garfield E. Citation analysis as a tool in journal evaluation[J]. Science, 1972, 178: 471-479.

[9] Lindholm-Romantschuk Y, Warner J. The role of monographsin scholarly communicationan empirical study of philosophy, sociology and economics[J]. Journal of Documentation, 1996, 52(4): 389-404.

[10] 苏新宁. 我国人文社会科学图书被引概况分析_基于CSSCI数据库[J]. 东岳论丛, 2009, 30(7): 5-13.

[11] 张希, 鄂丽君. 图书情报学专著被引分析[J]. 情报科学, 2010, 28(9): 1354-1358.

[12] Hammarfelt B. Interdisciplinarity and the intellectual base of literature studies: citation analysis of highlycited monographs[J]. Scientometrics, 2011, 86(3): 705-725.

[13] 杨思洛, 王皓, 文庭孝. 基于引文分析的图书影响力研究——以图书情报领域为例[J]. 情报资料工作, 2010(1): 89-92.

[14] Torres-Salinas D, Robinson-García N, Cabezas-Clavijo Á, et al. Analyzing the citation characteristics of books: edited books, book series and publisher types in the book citation index[J]. Scientometrics, 2014, 98(3): 2113- 2127.

[15] Bertin M, Atanassova I, Gingras Y, et al. The invariant distribution of references in scientific articles[J]. Journal of the Association for Information Science and Technology, 2015, 67(1): 164-177.

[16] Hu Z G, Chen C M, Liu Z Y. Where are citations located in the body of scientific articles? A study of the distributions of citation locations[J]. Journal of Informetrics, 2013, 7: 887-896.

[17] 胡志刚. 全文引文分析方法与应用[D]. 大连: 大连理工大学, 2014.

[18] 胡志刚, 陈超美, 刘则渊, 等. 从基于引文到基于引用——一种统计引文总被引次数的新方法[J]. 图书情报工作, 2013, 57(21): 5-10.

[19] Singh M, Patidar V, Kumar S, et al. The role of citation context in predicting long-term citation profiles: An experimental study based on a massive bibliographic text dataset[C]//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. ACM, New York, NY, USA, 2015: 1271-1280.

[20] 卢超, 章成志. 基于引文内容的单篇学术论文参考文献网络结构研究[J]. 现代图书情报技术, 2014(10): 33-41.

[21] 许秀江. 学术专著与学术论文的区别[EB/OL]. [2016/05/01]. http: //blog. sciencenet. cn/blog-667021-522174. html.


（责任编辑 车 尧）