

# 基于中国数据的国际管理研究学术进展<sup>①</sup>

## ——基于 UTD-24 和 FT-45 的可视化分析

陈宇新<sup>1</sup>, 张喆<sup>2\*</sup>, 杨涵方<sup>3</sup>

(1. 上海纽约大学商学院, 上海 200122; 2. 复旦大学管理学院, 上海 200433;  
3. 复旦大学计算机科学技术学院, 上海 201203)

**摘要:** 本研究基于 2009 年—2015 年 UTD-24 和 FT-45 使用中国数据源的文献数据库, 从研究领域、研究作者和关键词角度分析了基于中国数据的国际管理研究学术进展的现状和趋势, 并进一步基于数据挖掘和可视化技术提出描述学术研究发展趋势的一系列挖掘方法和呈现模式, 分析了商业管理学术领域对于中国数据源的关注程度、关注点以及研究模式, 以期更好地理解近年国际管理理论研究中中国数据的影响力和研究发展图景。在研究领域层面, 本文侧重于在学科领域、期刊和时间维度上对基于中国数据的关注程度进行分析。在研究作者层面, 本文侧重于发现关注中国数据源的研究学者的来源, 发现不同来源的研究者对中国数据源的不同研究视角及其之间的合作偏好。在关键词分析层面, 本文旨在发现基于中国数据学术研究中的热点问题, 把握研究主题随着时间的演变趋势。此外, 本文还感兴趣于关键词内部之间的关联性, 本文通过关键词与地域、研究领域的关联分析探索发掘各个领域或者地域对于研究主题的研究偏好, 通过关键词内部的关联性分析获得并发掘出潜在的可以协同研究的主题。

**关键词:** 中国数据源; 研究领域; 研究学者; 关键词; 数据可视化

**中图分类号:** C931 **文献标识码:** A **文章编号:** 1007-9807(2018)11-0061-15

## 0 引言

在 2007 年爆发的全球性金融危机和 2009 年相继爆发的欧债危机之后, 中国等新兴经济体表现依然十分活跃。2010 年中国更超越日本成为全球第二大经济体, 因而近些年越来越多的研究学者都相继把目光转向中国<sup>[1]</sup>。根据市场研究公司 IDC 的预测, 到 2020 年数字寰宇 (Digital Universe) 的规模将达到 40ZB, 相当于每个人可以均摊到 5 200GB 以上, 并且按照每两年翻一番的速度在扩张, 而其中 62% 将来源于新兴市场, 21% 来自中国<sup>[2]</sup>。随着数据驱动管理决策的时代到

来, 中国的数据量爆炸式增长, 信息化的管理方式也使得中国数据的质量更具有商业和学术研究价值。与此同时, 中国也从顶层设计 (以国家自然科学基金委员会管理学部为代表的科研管理机构) 积极推动基于中国情景和中国数据的高质量学术管理理论研究的开展, 并逐渐得到国际学术同行的认可<sup>[3-4]</sup>。本文旨在探索 2009 年以来商业管理学术领域对于中国数据源的关注程度、关注点以及研究模式, 以期更好地理解近年国际管理理论研究中中国数据的影响力和研究发展图景。

借鉴于 Mela, Roos 和 Deng<sup>[5]</sup> 以及 Murthi,

① 收稿日期: 2016-06-30; 修订日期: 2018-05-04。

基金项目: 国家自然科学基金资助项目 (71672038; 71428004; 71332006; 71222104); 国家留学基金资助项目 (201706105007)。

通讯作者: 张喆 (1973—), 女, 辽宁锦州人, 博士, 副教授, Email: zhezhang@fudan.edu.cn

本文作者为共同第一作者, 对论文有同等贡献。作者由衷感谢国家自然科学基金委管理科学部冯芷艳老师的指导和帮助。感谢复旦大学管理学院数据驱动管理决策研究中心的支持。

Rao 和 Ratchford<sup>[6]</sup>的研究方式,本文基于整个管理科学研究领域、研究作者和关键词内部之间的联系展开分析。首先,通过研究领域分析试图描绘在国际管理学术研究中对于中国数据源关注领域的分布情况,而创新理论和知识的发现归功于研究学者,因此进一步分析研究学者的研究模式。最后,通过关键词分析发现对于基于中国数据的研究问题的理论创新的研究视角和关注点以及发展趋势。因此,在研究领域层面,本文侧重在学科领域、期刊和时间维度上对基于中国数据的论文数量和比重进行分析。在研究作者层面,本文侧重于发现关注中国数据源的研究学者的来源,发现不同来源的研究者对中国数据源的不同研究视角及其之间的合作偏好。在关键词层面,本文尝试基于聚类分析观察关键词的重要性及其变化。此外,本文还讨论了关键词内部之间的联系性以及和领域及地域的关联度。

进一步,本文对研究成果进行了可视化的呈现,以期使研究者和应用者能够进一步观察、模拟和计算,形成新的数据理解和数据发现,这也是本研究的重要目标之一。数据可视化是将数据和信息编码为图形中的可视特征,然后充分利用人眼的高带宽对图形中的大量信息进行快速观察、分析、理解,其主要目标是通过图形含义向用户清晰和高效地传递数据中包含的信息<sup>[7]</sup>。随着数据可视化需求的增加,大量的开源或者商业的可视化工具集相继出现,本文中使用了其中两个常用的可视化工具: R 语言和 D3( Data-Driven Documents)。R 是一种同时可用于统计运算和产生数据图形的脚本语言( www. r-project. org)<sup>[8]</sup>。D3( Data-Driven Documents) 是一个 JavaScript 库,是一种基于 Web 的交互式可视化工具<sup>[9]</sup>。本文包含的可视化方法有: 类条形图( 堆栈条形图、Cleveland 点图、极坐标堆叠条形图)、饼图( 层次饼图和简单饼图)、协同图、关联性图表( 二分图、桑基图、和弦图)、世界地图、文字云图,以及泡泡图。在数据的分析上,本文使用统计学的方法探索数据的特征分布及其随着时间的变化,运用关联分析、聚类分析等数据挖掘方法探究数据之间的内部关系,主要使用 python 和 SQL 进行统计和挖掘。

本研究的主要贡献在于: 在理论方面,从研究

领域、研究作者和关键词内部之间的联系角度分析了基于中国数据的国际管理研究学术进展的现状和趋势,呈现近年来商业管理学术领域对于中国数据源的关注程度、关注点以及研究模式; 在方法方面,基于数据挖掘和可视化技术提出描述学术研究发展趋势的一系列挖掘方法和呈现模式; 在实践意义上,为科研管理机构的制度设计、研究学者的理论洞察、创新管理思想的应用传播提供了理论诠释和策略指导。

## 1 数据源

用于分析中国数据源使用情况的文献都源于 UTD24( 美国德州大学达拉斯分校商学院研究排名前 24 期刊) 和 FT45( 英国《金融时报》商学院研究排名前 45 个期刊) 上合计共 47 个期刊的发表论文。UTD-24 和 FT-45 是目前全球顶级商学院用于衡量学术研究能力和水平的通用标准。本文统计了这些期刊中 2009 年到 2015 年份的所有使用了中国数据源的文章及其作者的信息,合计共 558 篇文章,源于 1 275 位不同的研究者。数据源统计信息如表 1 所示。本文的所有分析都基于该数据源进行,即 2009 年—2015 年所有 UTD-24 和 FT-45 使用中国数据源的文献数据库。

文献的收集和统计本文运用 Google Scholar 进行初步查询,并在各个期刊所在的数据库中下载初步查询获得的源文献。进一步结合定点搜索,在每篇文章中关于数据源的介绍中查看其使用的数据源是否是中国数据源。最后获得文献列表。本文的数据主要包括四个部分: 论文信息、作者信息、各期刊每年发表的论文总量、期刊和领域信息。其中论文总量同样运用了 Google Scholar 进行统计。期刊和领域信息根据 UTD-24 和 FT-45 发布的情况而获得,主要包括期刊名字及其所属领域这两个信息。根据 FT-45,本文将 47 本期刊分为了以下 11 个领域( 详见附录): 1) Accounting; 2) Entrepreneurship; 3) General Management; 4) Marketing; 5) Organizational Behaviors; 6) Economics; 7) Finance; 8) Human Resource Management; 9) Operations and IT; 10) Practitioners; 11) Others。

论文信息包括: 论文题目、研究作者( 每个作

者包含一个作者 ID (用于与作者列表对应)、发表年份、关键字、数据源、数据源类型。本文将数据源类型分为两种: 1) 机构或企业统计的数据库中的数据(如 WIND、CCER 中国经济金融数据库、中国进出口数据、企业数据等); 2) 调研数据(survey data)。

作者信息则包括名字、学校、学校所在城市、州(省)、国家、其区域或者国家的 ISO 编码, 该城市的纬度及经度(其中后面三项属于绘制地图所需要的信息)。对于每位作者, 本文根据其所在学校和名字, 部分参考其在网络上的个人主页进行查重。而作者的学校则主要参考论文上提供的作者信息。若是包含多个学校的作者, 本文首先根据其办公地址和邮件选择其主要的学校进行录入; 若是无相关信息, 本文则选择其信息中第一所学校作为其学校信息。由于不同国家发表数量的较大差异和国家面积的差异, 若是单纯以国家为粒度则无法说明研究大国各个地区的发表差异。若是以城市作为划分, 对于一些体量较小的国家则划分过细, 会使得分析数据的长尾过长, 而且不便于读者理解。因此发表论文较多的国家先进行省(州)的拆分再细分到城市, 拆分的国家包括: 中国、美国、加拿大、澳大利亚、英国、印度。其中英国的拆分仅包括英格兰和苏格兰。

表 1 数据源统计  
Table 1 Data sources

统计年份	2009 年—2015 年
文章总数	558
数据源为数据库或者开源数据的文章数目	227
数据源为调研数据	331
作者人数	1 275
作者出现次数	1 630
来源国家和地区数目	36
来源研究机构数目	397
关键词数目	1 417
关键词出现次数	2 198
期刊数目	47
期刊领域分类	11

## 2 研究领域分布分析

本部分主要是分析基于中国数据源研究发表

在 2009 年—2015 年期间在各期刊和研究领域随着年份的变化情况。通过分析论文发表数量的绝对值(中国数据源论文发表篇数)及相对值(中国数据源论文该年发表篇数与其领域或期刊所有论文篇数的比值)的研究及对比, 可以观察到国际上商学院对于中国数据源重视和关注程度的不同, 从而帮助分析中国数据源学术研究最有价值的领域, 以及发掘尚未得到充分利用的数据源领域。

本文从三个方面进行对比分析: 研究领域的对比, 学术期刊的对比, 发表年份的对比。其中前两个部分由于各个领域和各个期刊每年发表文章数量很不相同, 所以本文分别对比了发表数量的绝对数量和相对比例。具体而言, 论文相对比例等于使用中国数据源的论文绝对数量占各领域和各个期刊论文总量的比例。

对应于三种对比分析, 本文运用了三种可视化技术: 堆栈条形图(stacked bar), Cleveland 点图(Cleveland Point), 极坐标堆叠直方图(polar coordinate)。在研究领域分布中领域数目较少, 使用堆栈条形图清晰明了地可以看到不同领域的发表情况, 及每年发表所占的比重。Cleveland 点图也是一种类似堆栈条形图的表示方法, 对于期刊数量较多的来说比条形图更为简洁清晰。每个点的位置即表示某个期刊在该年发表的论文数的累积值。

### 2.1 基于研究领域的分布分析

#### 2.1.1 绝对分布分析

图 1 使用堆栈条形图对比了在 UTD-24 和 FT-45 的 11 个研究领域发表基于中国数据源的研究成果的情况。条形图从下到上分别用不同颜色不同长度表示各个年份论文发表情况。总体而言, 使用中国数据源在“General Management”领域发表的论文数量远远高于其它领域, 在 2009 年—2015 年期间共发表 156 篇。论文发表数量体现出研究学者们对于基于中国数据源的基础管理理论研究问题的关注, 也体现出国家自科管理学部针对工商管理的基础研究项目支持的成果。

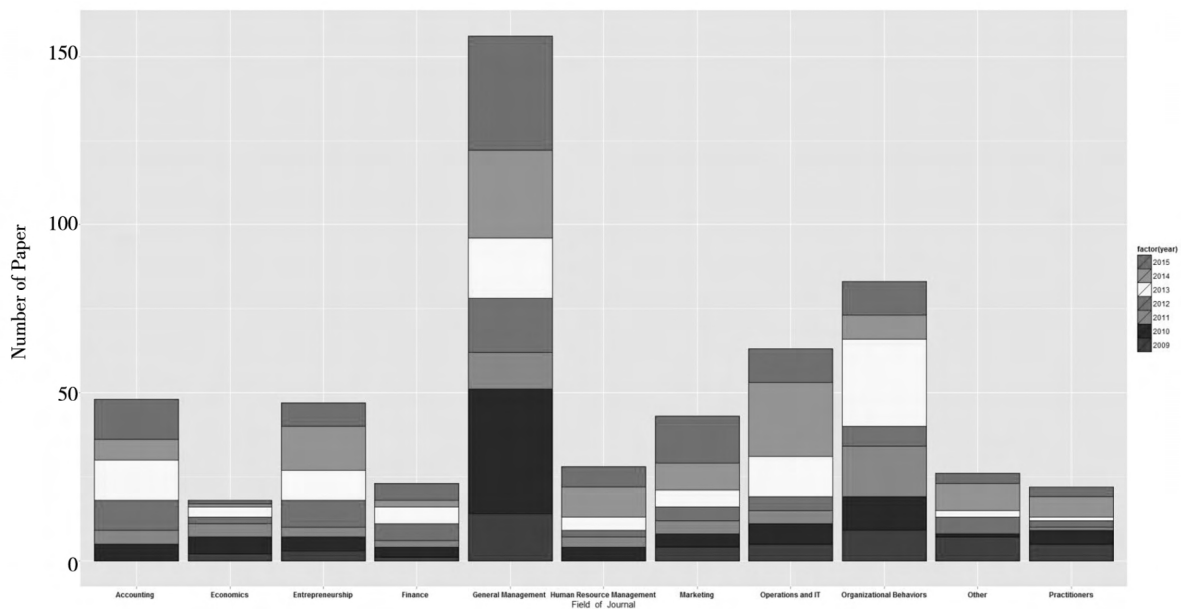


图1 基于中国数据源发表论文数量在各领域随年份累积分布直方图

Fig.1 The absolute cumulative distribution stacked bar chart of publications based on data sources from China

2.1.2 相对分布分析

进一步为了更客观表现各个研究领域对于中国数据源的关注程度,图2同样使用堆栈条形图对比了11个研究领域发表基于中国数据源的研究成果的相对比例

$$\text{某领域发表论文相对比例} = \sum_{i=2009}^{2015} R_i$$

$$\text{其中 } R_i = \frac{\text{该领域第 } i \text{ 年发表使用中国数据源的论文数}}{\text{该领域第 } i \text{ 年总发表论文数}}$$

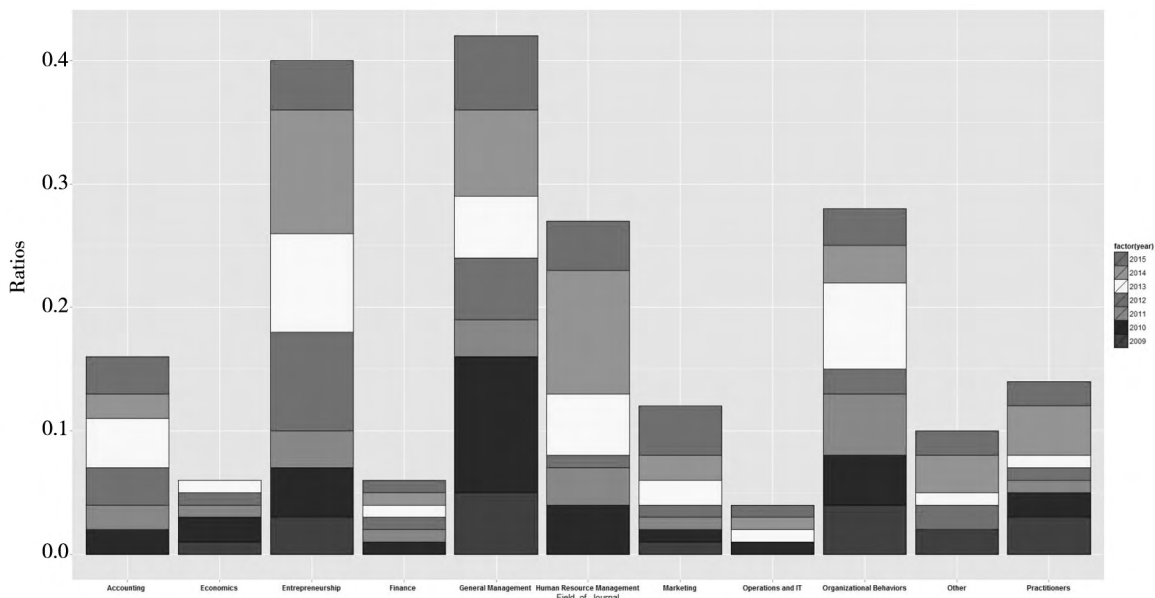


图2 基于中国数据源发表论文数量在各领域相对比率随年份累积分布直方图

Fig.2 The relative cumulative distribution stacked bar chart of publications based on data sources from China

从图2可以进一步确认,相比于其他10个研究领域,“General Management”无论从绝对分布还是相对分布都是非常关注基于中国数据源的研

究领域.也进一步表明,在基础管理理论实证研究领域对于中国研究情境的关注,其研究成果对国际管理理论发展的影响力相比于其他领域在学术

影响力和认可度上都高。而“Entrepreneurship”虽然发表基于中国数据的研究论文数量相比于其他领域较少,但发表的相对数量也是遥遥领先,仅次于“General Management”位居第二,也体现了在企业家能力研究方面中国数据源的优势和独特性。在研究领域方面,本文还发现在“Human Resource Management”领域的数据到目前为止,采用的数据全部为一手调研数据类型。这表明对于人力资源管理领域,目前实证研究中对企业人力资源系统信息数据和员工数据以及外部数据的运用有相当的潜力可以挖掘。

### 2.2 基于学术期刊的分布分析

#### 2.2.1 绝对分布分析

图 3 使用 Cleveland 点图对比了 47 个期刊发表使用中国数据源的发表文章数量情况。Cleveland 点图是一种与统计条形图类似的表示方式,以直线的长度表示因变量的统计图表,可竖排或者横排排列。图 3 中从左到右各点分别表示 2009 年—2015 年发表数量的堆叠,其中每两个点之间的线段长度则代表其右方所在点表示的年份发表的中国数据源论文数量。与基于研究领域的绝对分布分析类似,期刊文章的数量与该期刊对于中国数据源的关注程度和每年的发表数量共同相关。

总体而言 相比于其他期刊,“Journal of International Business Studies”、“Journal of Applied Psychology”和“Strategic Management Journal”在 2009 年—2015 期间共发表基于中国数据源的实证研究成果分别为 54 篇、46 篇和 41 篇,位居前三位。

#### 2.2.2 相对分布分析

图 4 使用 Cleveland 点图对比了 47 个期刊每年发表使用中国数据源论文数量相对值的不同情况。从中进一步验证期刊“Journal of International Business Studies”和“Journal of Applied Psychology”对于中国数据源的研究持续关注。而反之“Review of Financial Studies”,“Journal of Finance”和“INFORMS Journal on Computing”甚少关注中国数据源进行的研究,本文认为中国数据的质量以及研究学者关注的研究问题是导致问题的可能原因。

但同时本文也看到在如“Marketing Science”期刊上的可喜变化。随着营销学者对中国问题和中国数据的关注增加,近年来发表在“Marketing Science”的基于中国数据源的研究成果也在持续增长。这其中有海外华人学者、海归学者和本土学者的共同努力和合作结果。虽然数量还是较少,但相信在未来会有更多的中国研究成果出现在国际营销科学顶级期刊上。

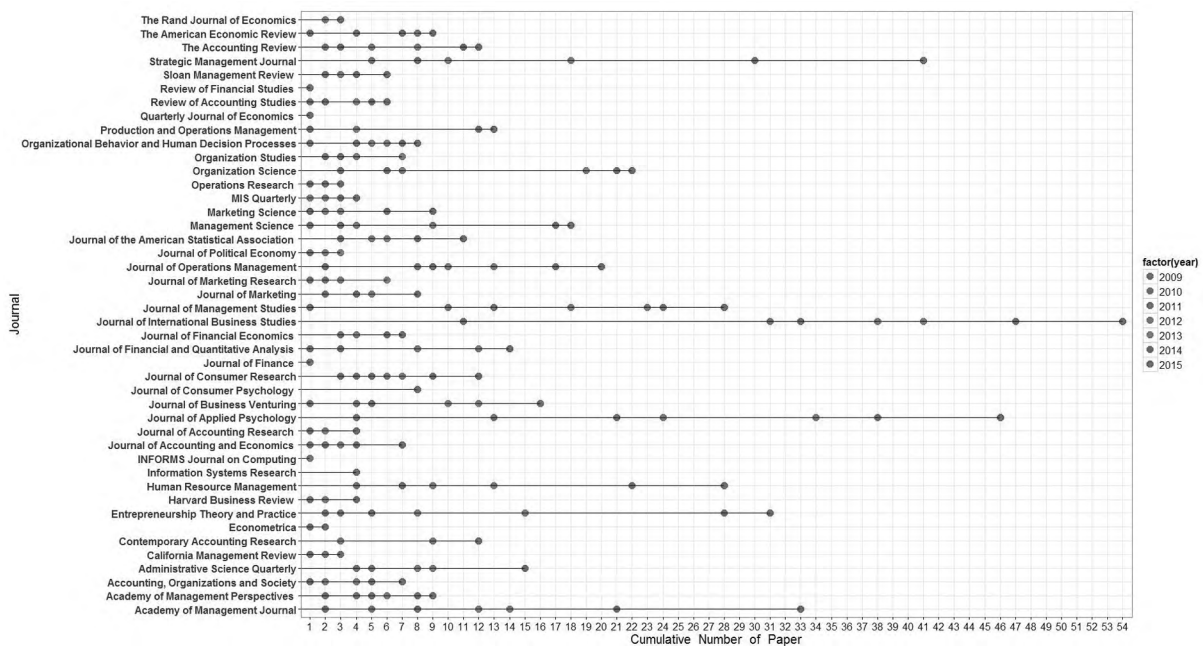


图 3 基于中国数据源发表文章数量在各期刊随年份变化 Cleveland 点图

Fig. 3 The absolute cumulative distribution Cleveland dot chart of publications in UTD-24 and FT-45 Journals based on data sources from China

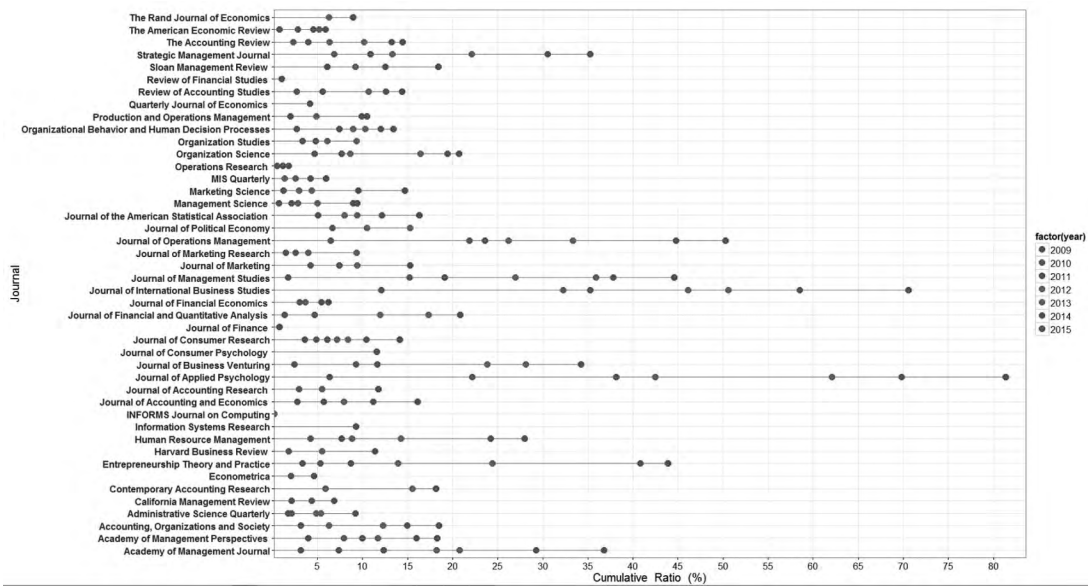


图4 基于中国数据源发表论文数量的相对比率在各期刊随年份变化 Cleveland 点图

Fig. 4 The relative cumulative distribution Cleveland dot chart of publications in UTD-24 and FT-45 Journals based on data sources from China

### 2.3 发表年份的分布分析

图5 通过极坐标直方图对比了2009年—2015年不同年份论文发表的数量变化情况,并通过各颜色显示该年份中各个领域所占的比重(领域所表示扇形所占半径长度越长,则表示该领域在当年份发表文章数量越高)。从图5和表2中可以看到,2010年开始,研究者对中国数据的关注有广泛增加,2011年及2012年则稍微冷却了两年

2013年研究量突增并在2014年和2015年保持基本稳定。本文认为,大数据作为互联网、移动计算、云计算之后IT又一次技术变革,正在近年深刻影响国家、企业、组织和个人决策过程和方式<sup>[3]</sup>。面向管理决策的大数据研究与应用逐步兴起,研究理念、思路、方法和学术路线和方向的探索已经开始全面展开。这是近年数据驱动管理决策研究快速增长的主要动因。

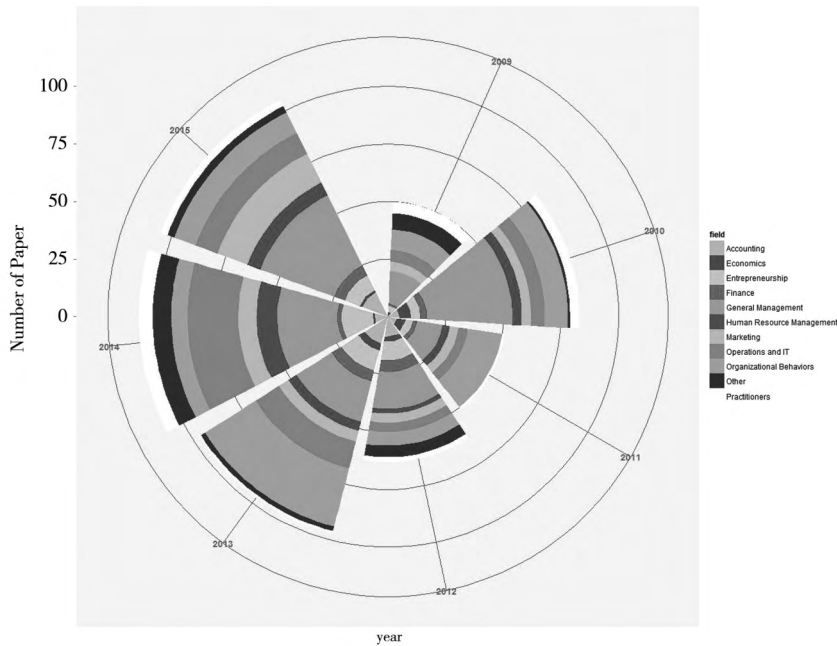


图5 基于中国数据源的发表论文数量随年份变化的极坐标直方图

Fig 5. The polar coordinate chart of publications based on data sources from China

表 2 2009 年—2015 年 UTD-24 和 FT-45 基于中国数据源的  
发表论文数量

Table 2 Publications from 2009 to 2015 in UTD-24 and FT-45  
Journals based on data sources from China

年份	基于中国数据源的发表论文数量
2009	50
2010	83
2011	51
2012	62
2013	96
2014	109
2015	104

### 3 研究者的分布与合作分析

本部分主要关注使用中国数据源的研究学者的分布、论文发表及相互合作的情况,使用按类划分的 Cleveland 点图、协同图、相关图、饼图及世界地图进行可视化展示。通过研究者的分析,可以发掘出对于中国数据源最感兴趣的研究学者和研究地域,这对于探讨科研组织的研究能力和研究取向很有效益,并且可以运用到其它领域或者研究汇总的探讨。通过研究合作的分析可以看出各个研究地域之间的相互交流合作状况。

#### 3.1 研究者来源分析

图 6 使用按类划分的 Cleveland 点图表示各个地域(根据体量进行不同粒度的划分:州、省级行政单位、地区、国家)作者数目的不同,并将各个地域根据国家和地区进行划分和聚类。按类划分的 Cleveland 点图类似于条形图,运用类似于火柴头的点和线对应  $x$  轴坐标表示论文的数量,同时根据  $y$  轴离散量的类进行排序和划分。从图 6 中可以直观地观察到在每个类内按照作者数量从多到少依次排列。研究者来源地统计了 1 275 个作者前 30 位(TOP30)来源地域的分布情况。来源地域统计关注来源地的多样性和每个地域作者的多元性,所以是关注的作者本身而非作者的发表数量,每个研究者不论发表数量多少都在统计中代表一个单位量。

研究者来源地域的排名中可以看到:①每个地域对于中国数据源的关注和研究人数;②各个地域的商业管理学术研究人才的储备水平;③各

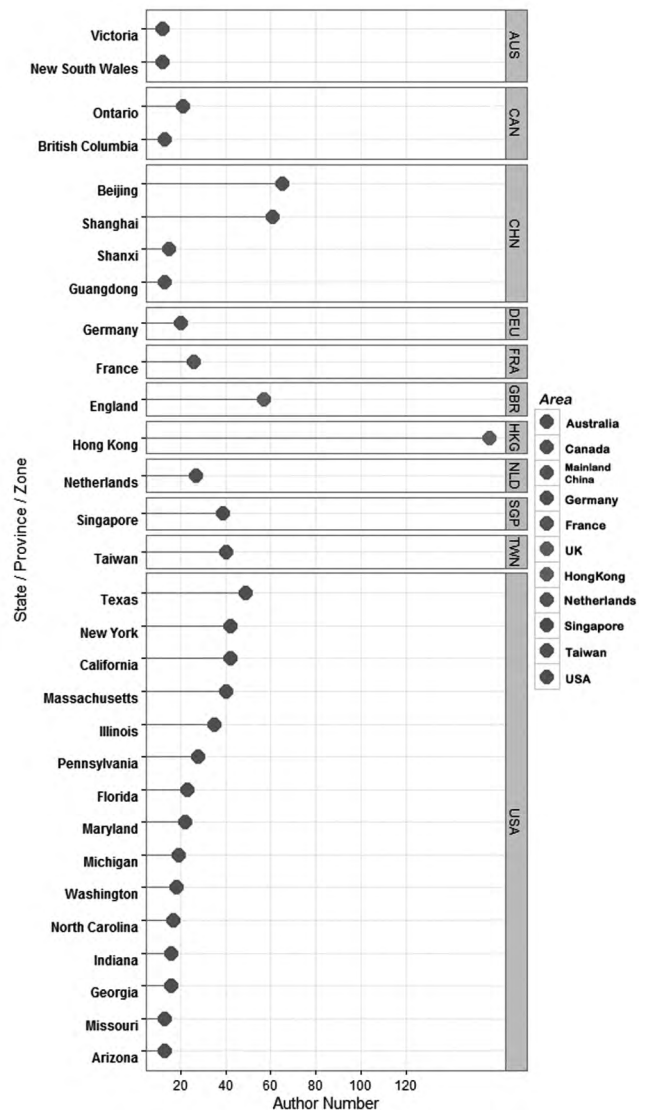


图 6 研究者数量按地域划分的 Cleveland 点图

Fig. 6 The Cleveland dot chart of researchers according to area

个国家或者地区中对于中国问题聚集关注的地域数量。分析发现,中国(包括香港和台湾)和美国都是研究者的主要来源地,而澳大利亚、加拿大、英国、新加坡等也有对于中国数据源聚集关注的地域。其中香港研究者最为集中且遥遥领先,高于排名第二在北京地区的研究人员数量两倍以上。而美国分布则更为广泛,在 TOP30 的研究人员聚集地域中占近一半的名额。这其中作为对自身国家的数据关注的自然性,所以中国研究论文位居榜首既是情理之中,也包含一部分对于中国研究水平的说明;而美国作为除中国(包括香港和台湾)之外发表论文最多的国家,也反映了美国在管理基础理论研究上的实力。

### 3.2 研究者论文发表数量分析

图7使用按类划分的Cleveland点图统计发表文章最多的前35位(TOP35)的研究者的发表文章数量,并按国家和地区归类.发表文章较多的作者主要来源于中国(包括香港和台湾)、美国和新加坡三个国家.其中中国和新加坡属于华人社会,对于中国数据源有较高的关切心理;而美国多元化的社会和较高的基础理论研究水平,使得美国当地有较多的关注中国数据的高水平研究学者.

新加坡三个国家.其中中国和新加坡属于华人社会,对于中国数据源有较高的关切心理;而美国多元化的社会和较高的基础理论研究水平,使得美国当地有较多的关注中国数据的高水平研究学者.

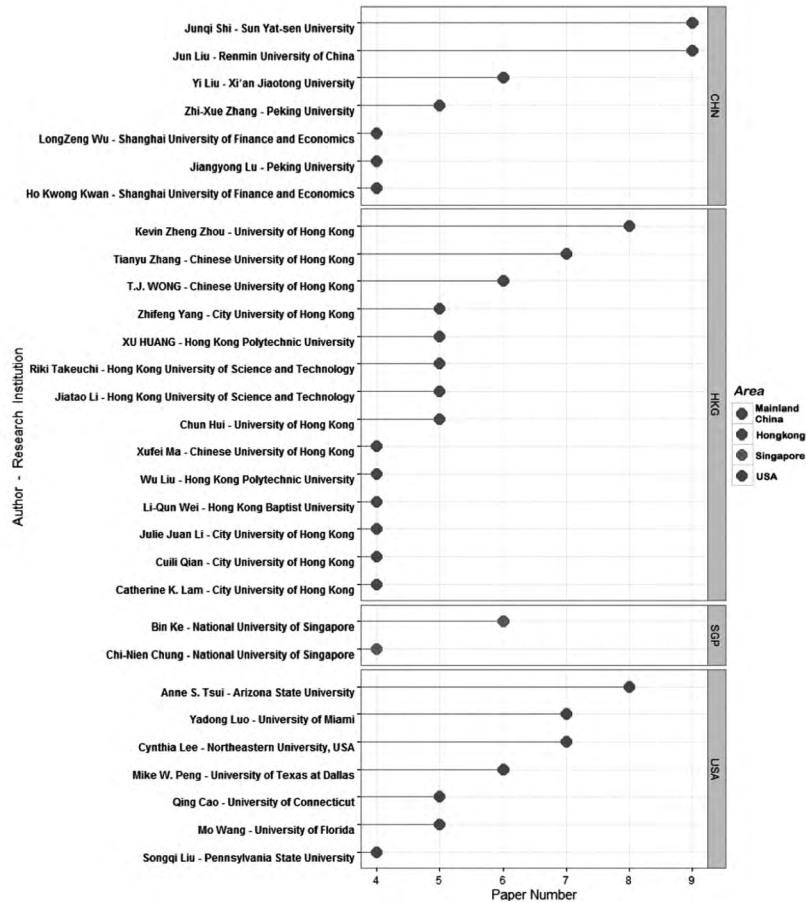


图7 研究者发表基于中国数据源论文数量的Cleveland点图

Fig.7 The Cleveland dot chart of researchers' publications according to data sources from China

### 3.3 合作分析

#### 3.3.1 研究者合作倾向分析

图8的饼图显示论文研究者合作方式的广泛性分析,对比倾向于仅校内合作撰写论文、同城市合作撰写论文和同国家合作撰写论文,跨国合作撰写论文所占的比率明显占主导形式.在发表论文的研究者中,与国外研究人员合作的数量占所有研究方式的大约59.2%,而与国内其他城市的研究人员合作的占比19.9%,选择与市内另一所学校以及自己学校合作的分别占比4.5%和16.4%,因而可以看出更广泛的国际合作是主要的合作倾向.主要原因本文认为基于中国数据的管理学术研究,需要管理、经济、统计、计算机等不

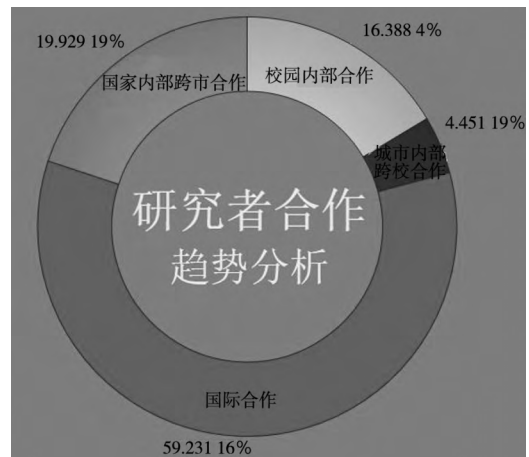


图8 研究者之间合作方式倾向分析饼图

Fig.8 The pie chart of cooperation tendency analysis among researchers



同学科的跨学科完成,中国学者在获取中国数据方面有良好资源,而海外学者在研究问题上有良好的洞察,在研究方法上有良好掌握,从而促进跨学科的国际合作的形成.

### 3.3.2 研究者地域的合作关系分析

图 9 通过协同图描述了各个国家或者地区之间合作的偏好图.图 9 中协同图中集中展示了前 10 位(TOP10) 论文发表合作数量的地域之间的合作关系.从中可以看到中美之间的研究学者具有较强的合作关系和合作倾向.深究其原因本文认为,在管理研究领域,目前美国是中国研究生和研究学者最大留学和访学目的国.通过中美学者的共同努力,联合培养了很多颇有建树的年轻学者,以此形成针对中国问题和对中国数据感兴趣的本土学者、中国海归学者,在美华人学者和美国本土学者的多元合作关系.

### 3.3.3 研究者合作和发表分布

图 10 采用地图的形式描述了各个国家发表论文的状况和合作关系.颜色标注关注的国家,以省(州)为单位,圆圈大小表示前 20 位(TOP20) 研究地域的

发表论文数量,黄色曲线表示前 150 位(TOP150) 合作关系对.按深浅排序,越深国家表示发表论文数量越多.从图 10 可以进一步验证中美在基于中国数据源的商业管理学术研究领域合作关系紧密.

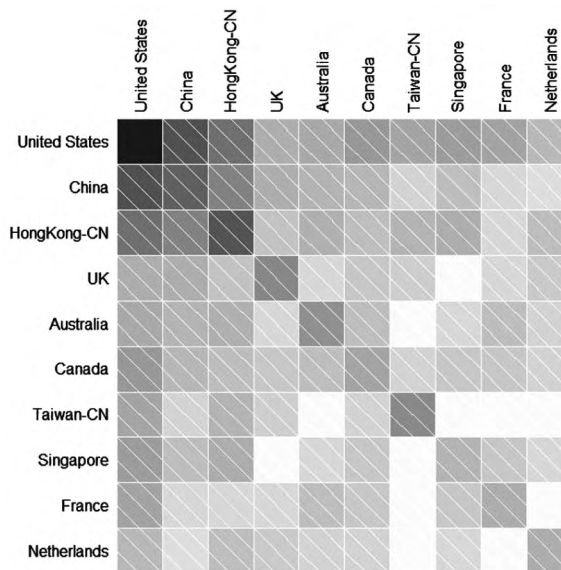


图 9 地域合作关系强弱协同图

Fig. 9 The collaboration diagrams of regional cooperation analysis

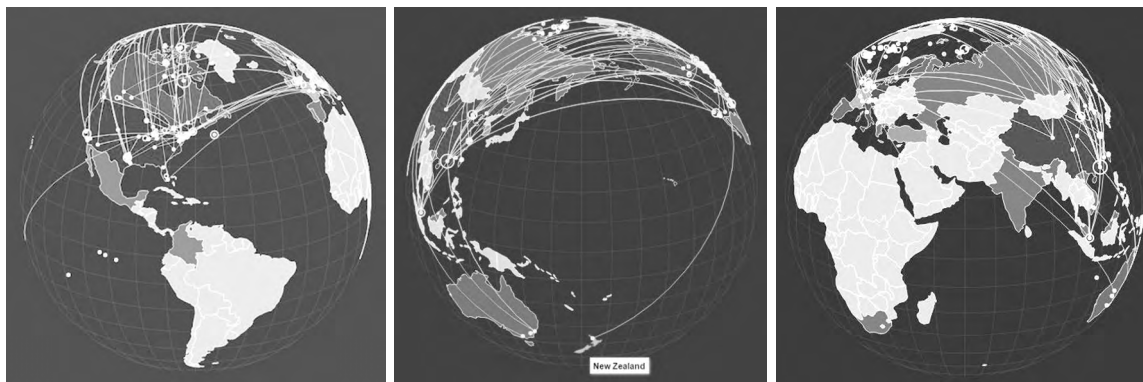


图 10 国家及区域合作地图(地球的三个截面地图)

Fig. 10 National and regional cooperation map

综上所述可以看出,由于商业管理领域实践和理论问题的多样性,本文发现大部分的研究成果都不是由完全同一个学校的作者完成的.说明研究者背景的多样性可以促使研究获得更为积极有效的成果.

## 4 关键词分析

对于关键词的统计分析,本文旨在发现基于中国数据学术研究中的热点问题,发现研究主题

随着时间的演变趋势.此外本文还感兴趣于关键词内部之间的关联性,本文希望通过关键词与地域、研究领域的关联分析探索发掘各个领域或者地域对于研究主题的研究偏好,通过关键词内部的关联性分析获得相似性的关键词,并发掘出潜在的可以协同研究的主题.

### 4.1 关键词文字云图

文字云图是关键词出现量统计结果的一种非常直观的视觉化描述方式,关键词其出现次数通过其大小进行描述,颜色随机变化便于区分.图

11 展示了基于中国数据源发表论文的关注的研究主题. 图中关键词大小与其出现次数  $N$  的平方根  $Sqrt(N)$  成正比, 对关键词出现次数  $N$  使用某些转换函数, 可以缩小或者增大不同关键词出现次数的方差, 可以更突出云图希望传达的主题. 例如, 在有少数词汇出现次数过大的情况下, 平方根函数或者对数函数可以更为明显地观察出现次数较多但与最大值有一定差距的关键词. 从图 11 本文可以看出目前发表在 UTD-24 和 FT-45 上基于中国数据源的研究主题明显与中国这个新兴市场的基础管理理论(如公司治理、文化、创新)相关.

#### 4.2 关键词随时间重要性变化

图 12 展示了词频高于 2 次关键词在 2009 年—2015 期间每年出现次数的变化. 圆圈大小与出现次数成正相关. 从图中可以看出有些研究主

题如公司治理、企业家精神、创新、信任等一直是基于中国数据源研究的关注主题, 而随着商业管理实践的推进, 如行为研究、创造力、文化等逐渐成为基于中国数据实证研究的关注对象.



图 11 关键词文字云图  
Fig. 11 Keywords cloud



图 12 高频关键词数量随时间变化点阵图  
Fig. 12 The bitmap of high-frequency keywords over time

### 4.3 主要关键词区域及领域分布

图 13 使用桑基图 (Sankey Diagram) 来描述 10 个主要的关键词在各个领域的分布情况, 以及研究这些关键词的国家和地区. 桑基图是流量图的变形, 其延伸的分支宽度对应数据流量大小. 在图 13 的桑基图里, 本文通过两个变量之间的连线宽度表示它们之间的关系紧密程度, 比如领域与关键词的宽度表述在某个领域关于该关键词文章数量多少, 国家或者地区与关键词连线的宽度表述该国家的研究人员发表和该关键词有关的文章数量多少. 相似的, 每个变量的宽度表示和该变量相关的文章数量多少.

本文研究发现: 每个地域关注的话题存在差异. 比如关键词人力资源管理 (HRM), 美国以外的国家对其使用中国数据源的实证研究较少并且基本都使用的是调研数据. 又例如, 对于跨文化的管理 (cross-culture management) 而言, 本文可以发现美国和新加坡比较重视这个研究主题, 这与其社会的多元化结构有关, 尤其华人在其社会中扮演重要角色.

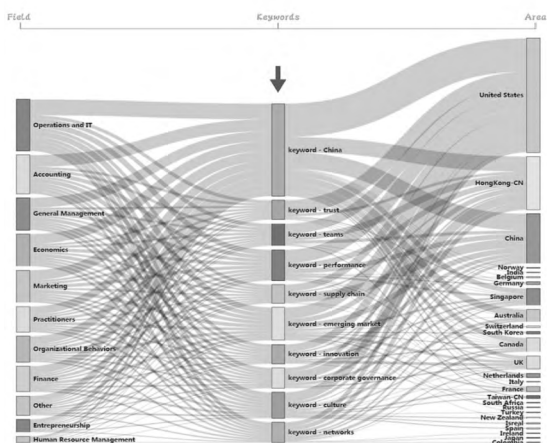


图 13 高频关键词与地域和研究领域相关性桑基图

Fig. 13 The Sankey diagram of correlation among high-frequency keywords, region and research field

### 4.4 关键词关联性分析

和弦图 (chord diagram) 是一种显示数据之间相互关系的一种图形化方式, 图 14 中词频排名较前的关键词被排列在圆圈的周围, 通过连线表示它们之间的关联度 (两个关键词若在同一篇文章中出现过则表示存在关联度). 这个部分探索了研究主题之间的研究热度. 关联性大的关键词联系表示两个话题存在很高的共性和热度; 反之比

较小的表示相关性较小, 也显示其暗含着被挖掘探索的关联度. 研究发现, 在 UTD-24 和 FT-45 上基于中国数据源的商业管理实证研究中信任 (trust) 与关系 (guanxi)、团队 (teams) 与创造性 (creativity) 具有高关联性. 这其中也体现对其因果性的重要启示.

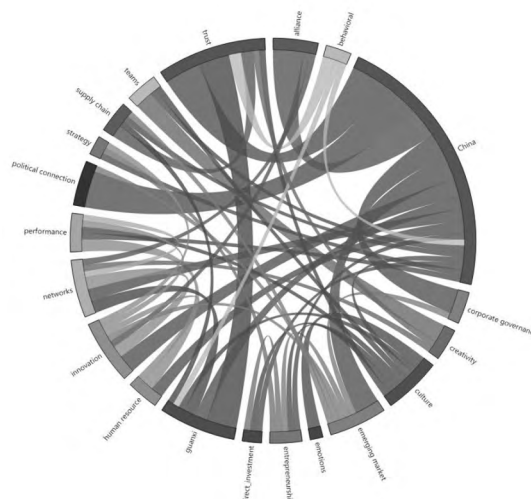


图 14 高频关键词之间相互关系和弦图

Fig. 14 The chord diagram of correlation between high-frequency keywords

### 4.5 关键词聚类分析

在基于图 15 关联性分析的基础上, 本文希望通过一种自动方式分析展示出词汇内部潜藏的相似含义. 提取每个关键词的特征主要考虑到以下两个因素: (1) 关联度——对于每一对关键词, 共同出现的文章篇数; (2) 关联总数——对每个关键词其它关键词与其建立关系的关联度总和. 根据这两个因素, 对于每一个关键词 (keyword), 建立一个维数为关键词总数量  $keywords\_length$  的特征向量  $rela\_vector$ , 对于任意一个 keyword 来说,  $rela\_vector [i]$  表示该 keyword 与第  $i$  个关键词之间的关联强度:

关联强度 =

$$\frac{keyword \text{ 与第 } i \text{ 个关键词的关联度}}{\sqrt{(keyword \text{ 的关联总数} * \text{第 } i \text{ 个关键词的关联总数})}}$$

特别的, 本文设置每个关键词与其自己本身的关联强度等于 1. 在获得每一个关键词的特征向量之后, 对所有的特征进行 k-means 聚类, 将所有关键词自动地进行分类 (本文对于所有 1 417 个关键词的分别提取出 1 417 维的特征, 聚拢到



发现,基于中国数据源的研究主题明显与中国这个新兴市场的基础管理理论(如公司治理、文化、创新)相关。(2)通过点阵图,本文看到有些研究主题如公司治理、企业家精神、创新、信任等一直是基于中国数据源研究的关注主题,而随着商业管理实践的推进,如行为研究、创造力、文化等逐渐成为基于中国数据实证研究的关注对象。(3)从关键词的区域和研究领域而言,使用桑基图,本文研究发现:每个地域关注的话题存在差异。(4)通过和弦图本文可以探索研究主题之间的研究热度。(5)通过关键词聚类分析,本文可以发现潜在的研究主题。

## 5.2 理论贡献

针对在 UTD-24 和 FT-45 中基于中国数据源的学术文献研究,更好地理解近年国际管理理论研究中中国数据的影响力和研究发展图景。可以看到,国际商业管理学术领域对“中国管理,中国数据”研究正日渐重视,但在广度和深度上都仍有很大的发展空间。同时,本研究在探索应用数据可视化方法研究学术领域的发展方面进行了有益的尝试,相信随着技术的进步和数据的丰富会有更多的研究认知被发现和被呈现,以此来共同来推进中国的基于数据驱动管理决策的学术研究的未来。

## 5.3 实践意义

本文的研究发现和研究成果,一方面可以为国家、地方以及高校科研院所科研管理部门的政策制定提供有益的建议,提高制度设计的有效性,促进在国际学术研究中形成基于数据驱动管理决策的中国方案,并对国际管理学术研究的商业实

践产生积极影响;另一方面也为管理领域的研究学者提供了未来的研究趋势和研究方向,以及研究发表路径的选择依据,并且在合作研究的形成方面获取更多的机会。此外,本研究为管理学术研究成果的商业化应用和传播机构提供了观察中国管理实践崭新视角的学术背景和商业传播洞察。

## 5.4 研究局限和未来研究方向

本研究还存在着研究局限和不足。首先,本研究基于 2009 年—2015 年 UTD-24 和 FT-45 使用中国数据源的文献数据库,从研究领域、研究作者和关键词角度运用统计分析和数据挖掘方法对基于中国数据的国际管理研究学术进展的现状和趋势进行了描述性分析。可以考虑对各种现状产生的原因及理论暗示进行深入分析,鼓励构建理论模型进行理论验证。比如可以借鉴 Jia, You 和 Du<sup>[1]</sup>的基于上下文模型(context-emic model)通过关键词文本挖掘分析基于中国数据研究的理论贡献。其次,本研究关注在 2009 年之后中国数据和中国方案在国际上愈发受到重视的研究区间。进一步建议可以考虑增加和延长时间跨度,将基于中国数据的管理决策研究的动态演变过程进行充分体现。并且进一步考虑以发表论文使用的数据年份为时间维度,进一步更好地理解国际管理理论研究中中国数据源的研究价值。最后,本研究鼓励在未来的研究中针对具体细分研究领域(如基础管理(General Management)、营销(Marketing)<sup>[10]</sup>、组织行为(Organizational Behaviors)、经济(Economics)、金融(Finance)<sup>[11]</sup>等)或者具体期刊进行更细粒度的分析和洞察。

## 参考文献:

- [1]Jia L, You S, Du Y. Chinese context and theoretical contributions to management and organization research: A three-decade review[J]. *Management and Organization Review*, 2012, 8(1): 173-209.
- [2]IDC. 第 6 次数字宇宙研究报告[R]. <http://www.idc.com/>, 2012.  
IDC. Digital universe[M/OL]. <http://www.idc.com/>, 2012. (in Chinese)
- [3]冯芷艳,郭迅华,曾大军,等. 大数据背景下商务管理研究若干前沿课题[J]. *管理科学学报*, 2013, 16(1): 1-9.  
Feng Zhiyan, Guo Xunhua, Zeng Dajun, et al. On the research frontiers of business management in the context of big data [J]. *Journal of Management Sciences in China*, 2013, 16(1): 1-9. (in Chinese)
- [4]徐宗本,冯芷艳,郭迅华,等. 大数据驱动的管理与决策前沿课题[J]. *管理世界*, 2014, 11: 158-163.  
Xu Zongben, Feng Zhiyan, Guo Xunhua, et al. Big data driven management and decision-making frontier issues[J]. *Management World*. 2014, 11: 158-163. (in Chinese)
- [5]Mela C F, Roos J, Deng Y. A keyword history of marketing science[J]. *Marketing Science*, 2013, 32(1): 8-18.

- [6]Murthi B P S , Rao R C , Ratchford B T. A descriptive analysis of publications in marketing science over its history [J]. *Marketing Science* , 2015 , 34( 6) : 779 – 786.
- [7] Friedman V. Data visualization and infographics [M/OL]. *Smashing Magazine* , <https://www.smashingmagazine.com/>. 2008.
- [8]R Core Team. The R project for statistical computing [M/OL]. <https://www.r-project.org/>. 2018
- [9]Bostock M , Heer J , Ogievetsky V. D3: Data-driven documents [J]. *IEEE Transactions on Visualization and Computer Graphics* , 2011 , 17( 12) : 2301 – 2309.
- [10]Sudhir K. Marketing Science: Frontiers [J]. *Marketing Science* , 2018 , 37( 1) : 1 – 4.
- [11]冯 博 , 叶绮文 , 陈冬宇. P2P 网络借贷研究进展及中国问题研究展望 [J]. *管理科学学报* , 2017 , 20( 4) : 113 – 126.
- Feng B , Ye Q W , Chen D Y. Review on P2P online lending and new research opportunities for China 's case [J]. *Journal of Management Sciences in China* , 2017 , 20( 4) : 113 – 126. ( in Chinese)

## The recent developments in international management research with China data source: An analysis based on publications in UTD-24 and FT-45 journals with data visualization

CHEN Yu-xin<sup>1</sup> , ZHANG Zhe<sup>2\*</sup> , YANG Han-fang<sup>3</sup>

1. Business Division , New York University Shanghai , Shanghai 200122 , China;
2. School of Management , Fudan University , Shanghai 200433 , China;
3. School of Computer Science and Technology , Fudan University , Shanghai 201203 , China

**Abstract:** This research examines recent developments and trends in international academic research on business and management that utilizes data from China. Information collected are from papers published from 2009 to 2015 in the 24 journals used by the University of Texas at Dallas Business School Research Rank ( UTD-24) and the 45 journals used by Financial Times Research Rank ( FT-45) . After identifying all papers using data source from China in those journals , the paper further examines the variations of China data based publications across disciplines and journals and the dynamics of those variations over time , the patterns of research collaborations among individual researchers and research areas with publications using China data source , and the evolution and associations of research topics as represented by the keywords used in the publications with China data source. Leveraging the latest developments in data science , advanced data mining and visualization techniques are adopted to aid our analysis and illustrate our findings. Our research shed light on the global impacts of data from China on cutting-edge academic research in business and management , as well as the trends in research developments with China data source.

**Key words:** China data source; research domain; authors; keywords; data visualization

附录: 期刊列表及其所在领域( UTD-24 和 FT-45)

	期刊名字	数据库	领域	说明
1	Accounting , Organizations and Society	Elsevier	Accounting	FT
2	Journal of Accounting and Economics	Elsevier	Accounting	UTD & FT
3	Journal of Accounting Research	University of Chicago	Accounting	UTD & FT
4	The Accounting Review	American Accounting Association	Accounting	UTD & FT
5	Review of Accounting Studies	Springer	Accounting	FT
6	Contemporary Accounting Research	Wiley	Accounting	FT
7	Entrepreneurship Theory and Practice	Baylor University , Waco , Texas	Entrepreneurship	FT
8	Journal of Business Venturing	Elsevier	Entrepreneurship	FT
9	Academy of Management Journal	Academy of Management , Ada , Ohio	General Management	UTD & FT
10	Academy of Management Review	Academy of Management	General Management	UTD & FT
11	Journal of International Business Studies	Academy of International Business	General Management	UTD & FT
12	Strategic Management Journal	John Wiley and Sons	General Management	UTD & FT
13	Journal of Management Studies	-	General Management	FT
14	Journal of Marketing	American Marketing Association	Marketing	UTD & FT
15	Journal of Marketing Research	American Marketing Association	Marketing	UTD & FT
16	Journal of Consumer Research	University of Chicago	Marketing	UTD & FT
17	Marketing Science	Informa	Marketing	UTD & FT
18	Journal of Consumer Psychology	-	Marketing	FT
19	Journal of Applied Psychology	American Psychological Association	Organizational Behaviors	FT
20	Organization Science	Informa	Organizational Behaviors	UTD & FT
21	Organizational Behavior and Human Decision Processes	Academic Press	Organizational Behaviors	FT
22	Organization Studies	SAGE	Organizational Behaviors	FT
23	Econometrica	Econometric Society , University of Chicago	Economics	FT
24	Journal of Political Economy	University of Chicago	Economics	FT
25	The American Economic Review	American Economic Association , Nashville	Economics	FT
26	The Rand Journal of Economics	The Rand Corporation	Economics	FT
27	Quarterly Journal of Economics	MIT	Economics	FT
28	Journal of Finance	Blackwell	Finance	UTD & FT
29	Journal of Financial and Quantitative Analysis	-	Finance	FT
30	Journal of Financial Economics	Elsevier	Finance	UTD & FT
31	Review of Financial Studies	Oxford University Press	Finance	UTD & FT
32	Human Resource Management	John Wiley and Sons	Human Resource Management	FT
33	Information Systems Research	Informa	Operations and IT	UTD & FT
34	Journal of Operations Management	Elsevier	Operations and IT	UTD & FT
35	MIS Quarterly	Management Information Systems Research Centre , University of Minnesota	Operations and IT	UTD & FT
36	Management Science	Informa	Operations and IT	UTD & FT
37	Operations Research	Informa	Operations and IT	UTD & FT
38	Production and Operations Management	POMS	Operations and IT	UTD & FT
39	Academy of Management Perspectives	AMP	Practitioners	FT
40	California Management Review	UC Berkeley	Practitioners	FT
41	Harvard Business Review	Harvard Business School Publishing	Practitioners	FT
42	Sloan Management Review	MIT	Practitioners	FT
43	Administrative Science Quarterly	Cornell University	Other	UTD & FT
44	Journal of the American Statistical Association	American Statistical Association	other	FT
45	The Journal of Business Ethics	Kluwer Academic	other	FT
46	INFORMS Journal on Computing	INFORMS	Operations and IT	UTD
47	Manufacturing and Service Operations Management	-	Operations and IT	UTD