

# 基于 SCOPUS 的数据科学发展现状与趋势分析\*

曹瑞雪<sup>1</sup> 马英红<sup>2</sup> 李海洋<sup>1</sup> 于青林<sup>1,3</sup>

(1. 西安工程大学理学院 西安 710048)(2. 山东师范大学管理学院 济南 250014)

(3. 加拿大汤姆森河大学理学院 加拿大)

**摘要** 人们的日常生活越来越多地依赖于数据,从 GPS 定位到网上购物,研究和探索数据的工作变得十分重要。数据科学作为以分析数据为中心的学科分支,得到了科技界更多的重视,同时探索数据科学会为该学科发展的轨迹以及未来的发展提供更有用的信息和指导作用。通过检索 SCOPUS 中有关数据科学的论文,采用文献计量分析法从论文产出量、论文来源国家以及论文来源机构等角度对文献进行分析研究,运用统计学的方法和技巧,借助 R 语言、SAS 软件,对数据科学相关文献的分布及发展进行多方面的研究。解析该领域的研究现状和发展趋势,以期对数据科学研究的深入开展和未来大数据的发展提供参考信息。

**关键词** 数据科学;文献计量学;趋势;拟合;预测;聚类分析

**中图分类号** G72 **DOI:**10.3969/j.issn.1672-9722.2019.04.001

## A Study of Status and Trends of Data Science Discipline Based on SCOPUS

CAO Ruixue<sup>1</sup> MA Yinghong<sup>2</sup> LI Haiyang<sup>1</sup> YU Qinglin<sup>1,3</sup>

(1. School of Science, Xi'an Polytechnic University, Xi'an 710048)

(2. School of Management, Shandong Normal University, Jinan 250014)

(3. School of Science, Thompson Rivers University, Canada)

**Abstract** Daily life of the modern society more and more relies on data, from GPS positioning to online shopping. Therefore, exploration and understanding of data usage become an important research topic. As a disciplines centered on analyzing data, data science has received more attention from the scientific and technological communities. In the meantime, exploring the growing patterns of data science will provide more insights and guidance for the development of the discipline and its future growths. Based on searching results of data science publications in SCOPUS database, bibliometric methods are used to analyze publication outputs, countries of origins of publications, author institutions. Utilizing statistical methods and techniques, and deploying R language and SAS software, a variety of research on the distribution and the development of data science related publications are conducted. The study reveals the statuses and the developing trends of data science, and thus provides in-depth information for the future development of data science and big data.

**Key Words** data science, bibliometrics, trend, data correlation, forecasting, clustering

**Class Number** G72

### 1 引言

学科文献的数量变化反映学科知识量的变化情况,学科知识量的增长及其规律与文献的增长及

其规律是紧密相连的,因此学科文献是衡量学科知识量的重要尺度之一<sup>[1]</sup>。文献计量学是以文献量、作者数、词汇数等数据来研究学科知识量变化的学科。学科文献数据可以从历史性和连续性的数

\* 收稿日期:2018年10月11日,修回日期:2018年11月24日

基金项目:国家自然科学基金面上项目(编号:71471106,11271297);陕西省海外百人计划资助。

作者简介:曹瑞雪,女,硕士,研究方向:数据分析。马英红,女,博士,教授,研究方向:复杂网络。李海洋,男,博士,教授,研究方向:稀疏信息处理、量子逻辑、格上拓扑学等。于青林,男,博士,教授,研究方向:图论,网络优化,数学规划,数据分析。

据中找出学科发展规律<sup>[2]</sup>。因此主要采用文献计量分析体系,运用统计学的方法和技巧,使数据科学学科发展的现状与趋势更清晰明了。2004年,刘婧<sup>[3]</sup>通过统计中国知网数据库与中文科技期刊数据库中洛特卡定律和普赖斯定律相关的文献,经过时间、期刊、作者等多方面的整理分析,获得了相关文献的发展历程与变化。2010年,邱均平和杨瑞仙<sup>[4]</sup>根据ESI和WOS数据库,以临床医学领域热门论文为数据,从国家分布、机构分布、期刊分布等方面研究了临床医学的发展。2016年,Floriana Gargiulo等<sup>[5]</sup>基于Mathematics Genealogy Project数据库的博士数据从国家与学科两方面对数学近代史进行了研究。运用复杂网络工具对国家进行聚类分析,并对机构名次折线分布,博士国家三维分布,学科交叉程度分布,学者家族比例分布等进行了系统分析,并将分析结果与数学发展的历史与现状相比较。2017年,赵蓉英、魏明坤<sup>[6]</sup>基于citespace软件对国际数据科学进行了时间维度分析。通过文献计量分析以及数据科学的发展演进图谱研究,将数据科学分为了以数据储存为中心和以数据处理为中心的两个发展阶段,并对两个阶段的发展进行了细致的探讨。“data science”是一个新的学科领域,通过对“data science”相关文献的研究,从地域、机构等方面介绍了“data science”发展过程及发展现状,并对于将来的发展趋势<sup>[7-8]</sup>作出预测。

## 2 数据来源以及研究方法

本文所选取的数据源是SCOPUS数据库,该数据库是全世界最大的摘要和引文数据库,在文献检索方面具有很高的权威性。针对“data science”的论文产出量、论文来源以及论文的年度变化等情况进行了分析和探讨。在论文产出方面,起初检索与“论文标题,摘要,关键字”相关的“data science”论文,由于数据库2002年-2004年美国与“data science”相关的论文数据从11291篇突降到了3323篇,导致总的数据从33706篇突降到8472篇,因为数据库中数据的这种无故跌落现象未有合理的解释,所以在论文产出规律分析中,检索了SCOPUS数据库中1980年-2016年“任意字段”与“data science”相关的论文记录,共获得文献8627275篇论文数据。在论文来源国家分布方面,对SCOPUS中1980年-2016年在“标题,摘要,关键字”中含有“data science”的论文,按全球发文量90%的39个国家的发文量进行了检索。机构分布方面,获取了

SCOPUS数据库中1980年-2016年每年发文量排名前30名的机构。

本文的主要理论基于文献计量学五大定律以及系统聚类的四类方法。文献计量学的五大定律包括齐普夫定律、洛特卡定律<sup>[9]</sup>、布拉德福定律、加菲尔德定律以及普赖斯定律<sup>[10]</sup>。系统聚类的四类基本方法包括最短(长)距离法、类平均法、中间距离法以及重心法。这些基本定律和聚类方法是科学计量学、文献计量学研究中的基本知识。本文基于上述方法,针对“data science”学科中,论文的产出量、论文来源等分类形式,对现状进行了分析和探讨,并对“data science”学科的发展趋势进行了预测。

研究中,使用R语言<sup>[11]</sup>、SAS软件对数据科学从论文产出,来源国家分布,来源机构分布方面研究了相关文献的分布规律。论文产出方面,通过数据科学方法,将1980年-2016年的论文量检验了莱普斯的指数增长规律并运用SAS软件对五年内的数据发展进行了预测。来源国家分布方面,从数据科学的不同发展阶段出发,基于柯尔莫哥洛夫-斯米尔诺夫距离<sup>[12]</sup>对38个国家分别聚类,得出不同阶段这些国家在数据科学领域所属的类别。论文来源机构研究是基于SCOPUS网站1980年到2016年每年发文量前30的机构,通过计算相邻两年的机构变动程度,并通过统计学中的滑动平均处理,得到机构变动波动较大的两年,找出波动较大的机构,并根据机构排名变化找出不同时期在数据科学领域起带头作用的机构。通过以上统计分析得出数据科学论文在全球的发展规律。

## 3 论文产出量分析

利用文献计量学方法分析数据科学领域发展历程时,根据科技文献的增长及老化规律,对相关论文发文量进行年度统计分析,揭示该领域不同发展阶段的发展状况。并结合文献计量学的相关定律对数据科学的发展进行了预测。

文献计量学的这五大定律中,与相关的是普莱斯提出的科学文献指数增长规律。这里借助论文增长的指数规律,对数据科学1980年-2016年相关的论文进行数据统计,并用1980年-2014年的论文数据进行指数拟合,对2015年-2021年进行预测,其中用2015年与2016年的论文数据对数据预测的准确性进行验证。为直观地表达数据科学论文的变化趋势,以发文时间为横坐标,发文量为纵坐标,画出数据科学发文随时间变化散点图(如图1所

示)。根据数据科学文献的散点变化可以看出有相对明显的数据增量的是1996年和2004年,这两个增长点可以得到合理的解释,因为1996年的“data science”首次作为术语出现在会议标题中:“Data Science, classification, and related methods”<sup>[13]</sup>,2001年,美国计算机科学家 William S. Cleveland将“data science”作为一门独立学科介绍,在2004年数据科学这门学科开始被大家普遍接受并开始应用。根据图1数据科学文献的发展拟合曲线,可知数据科学论文的产出接近指数型增长,符合莱普斯指数增长规律并且曲线的拟合度较高( $R^2=0.9983$ )。所以可以通过SAS软件比较自信的做出2014年后七年的预测,用2015年-2016年的数据对论文的预测准确性进行验证。因为数据是非平 $b>0$ 稳时间序列,所以通过构建自回归滑动平均模型ARMA(1,1)对2014年后七年的论文量进行预测: $\frac{1+0.57133B}{1-0.85975B}\varepsilon_t$ ,其中 $B$ 为延迟算子,相当于把当前序列值的时间回拨了一个时刻, $\{\varepsilon_t\}$ 为随机干扰序列。通过图2对后七年的论文量拟合预测,预测到2015年和2016年的论文量分别为740192篇和796584篇,数据库中2015年和2016年的论文量为731294篇和775968篇,通过数据验证,预测准确性很高。所以可以相对保守估计出,在2016年之后未来发展中,五年内会发文4887767篇,近三年内数据会有32.1%的增长率。为了检验预测的可靠性,利用随机抽样抽取 Science, Nature, Lecture Notes In Computer Science (Including subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics) 三种杂志,得到1980年-2016年“data science”论文在这三种期刊总发文量中所占比例,通过平滑处理后得到数据科学在期刊发文量中的占比变化折线图。结果如图3所示,在三种杂志中,“data science”论文所占比例从1980年的0.0071逐渐上升至2016年的0.037。在此验证了上文对“data science”学科发文量趋势预测的正确性。

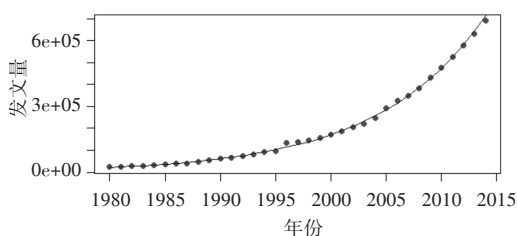


图1 1980年-2014年国际数据科学文献增长年度分布

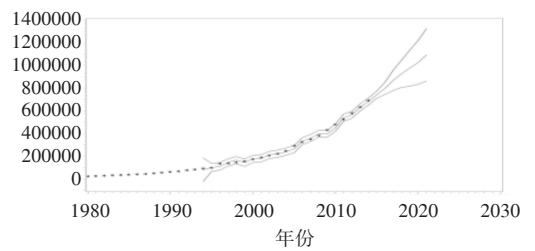


图2 数据科学2015年-2021年七年发文量增长预测

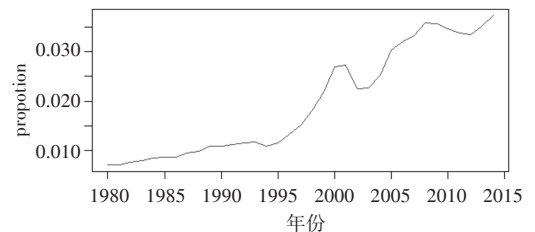


图3 数据科学在期刊发文量中占比曲线图

## 4 论文来源国分析

研究论文的来源国家分布,能帮助研究人员认清当前在某一领域较为进步的国家 and 地区,从而帮助研究人员认识本国该领域在国际上所处的位置,找到标杆国家,从而进行更有效的学习研究<sup>[14]</sup>。SCOPUS数据库中收纳了167个国家的发文量信息以及24905篇国家归属未知的文章。因为前39个国家的发文量达到总量的90%,所以只对前39个国家进行聚类分析。根据数据科学不同发展阶段进行聚类,从而探究数据科学在各个国家的发展情况。

获取占全球发文量90%的39个国家1980年-2016年发文量后,通过对数据标准化处理: $f_i(t) = \frac{N_i(t)}{N_j(t)}$ ,其中 $f_i(t)$ 为I国在 $t$ 时刻论文所占本国总发文量的比重, $N_i(t)$ 为 $t$ 时刻I国与数据科学相关的发文量, $N_j$ 表示I国从1980年-2016年的总发文量。标准化数据后基于柯尔莫哥洛夫-斯米尔诺夫距离 $D = \max|f(x) - g(x)|$ 求得距离矩阵 $D$ 。因为自变量是时间,所以距离矩阵 $D$ 是基于时间求得的两两国家论文数据随时间演变的最大值。获得距离矩阵后,结合系统聚类法将其聚类。对一般问题,各种聚类得出的结果通常都不一样。每一种聚类都代表某种合理但具有倾向的解释。但是,通过对1980年-2016年的国家进行聚类方法实验比较后,发现在的问题中,聚类的划分都趋向一致。说明利用的距离矩阵和系统聚类方法,结果比较稳定,有很好的可靠性。为了计算的方便,下面均选择系统聚类方法中的类平均法,分时间段对

国家数据进行聚类。

在1996年和2001年数据科学有两件标志性的事件,分别是第一次提出数据科学和数据科学成为一门独立学科的时间,因此以1996年和2001年为结点,将时间区域划分为1980年-1996年,1996年-2001年,2001年-2016年。针对数据科学三个不同发展阶段,对39个国家进行了聚类分析,使得三个时间段内论文变换走势接近的国家聚成一类。折线图可以显示随时间 $t$ 而变化的连续数据,可以看出在相同的时间间隔下,数据的趋势变化情况。所以将聚类和折线图结合对三个时间段每一类的国家论文趋势进行分析。

在1980年-1996年的聚类中,为了使国家间有相对的可比性,让分类尽量丰富并减少单个国家分一类的情况,通过实验比较,选择了将1980年-1996年的国家论文数据分为9类。通过对1980年-1996年,1980年-1996年,1980年-1996年的聚类图以及每一类的折线走势,可以直观地了解到早期各类国家的论文走势。从附录中1980年-1996年来源国聚类图以及各类别折线趋势图可以看出,早期各国家的数据科学论文比例都呈现出一定的波动,其中类5中的芬兰、捷克和类8的马来西亚呈现较多的数据波动,说明早期数据科学的发展中,相比其他国家,这些国家的数据科学发展还不稳定。从总的折线图趋势可以看出所有国家的论文数据都在1994年或1995年开始增长,说明数据科学在1996年第一次被提出前,很多国家对数据科学都已经开始发展,也为1996年数据科学的提出提供了基础。通过1980年-1996年的国家聚类结果看出,中国与韩国较为接近,并且与日本、加拿大等发达国家聚为一类,说明早期在数据科学提出时,这些国家对一个新领域的提出都有很好的敏感度。从数据科学各国家发文量占全球发文量所占比例来看,早期美国发文量所占比例远高于同时期的其他国家。

为了不使单个国家聚为一类的情况增多,例如新加坡不独自成为一类,使得各类别之间更有比较性,选择将国家聚为9个类别。可以通过附录中1996年-2001年论文来源国各类别折线图看出,1996年之后,论文来源国的各个类别走势接近,并且各国家的论文走势不再像1996年之前一样数据波动并且趋势各异。各国家的数据开始呈现出类似规律的变化,总体呈现出在1998年前的小幅度增长后,在1998年-1999年论文小幅度降低后开始快速上升。相比其他国家的数据科学发展,类6中

的马来西亚与中国在数据科学提出到成为独立学科这几年,发展起步较为缓慢。通过1996年-2001年的论文来源国聚类图得出,美国作为科技排名第一的国家自成一类。加拿大、德国、意大利、法国、日本等发达国家论文走势接近,聚为一类,说明这些发达国家早期对数据科学呈现出很强的学科灵敏度。中国与韩国、瑞士、英国、俄罗斯等国家,聚为一类,说明在数据科学刚提出的近几年,这些国家也开始投入研究并发展数据科学。

为了避免单一国家(比如美国)聚为一类,使得各类别内的国家更多,选择将2001年-2016年39个国家聚为7类。当类别从9类变为8类时,澳大利亚,南非两个国家的类别组合进日本、马来西亚等国的类7,当类别从8类变为7类时,美国由独立的类别加入进了原本新加坡与乌克兰的类3。通过类别的减少变换,可以找到各国家相对距离较近的类别,从而探索以往单独一类的国家所属类别的变化。通过图5的折线趋势图得出,数据科学在2001年正式成为一门独立学科后,各国家的论文比例普遍在2002年-2004年降低,说明随着数据科学成为一门独立学科后,数据科学论文的定义与要求都有了更全面规范的要求。2004年以后,各国家的论文比例普遍增长,说明随着数据科学成为独立学科并规范要求后,该学科开始呈现相对稳定的增长。通过图4的国家聚类图得出,美国作为科技最发达的国家,随着类别减少,开始与新加坡、乌克兰聚为一类。奥地利、新西兰,意大利、比利时、加拿大等发达国家聚为一类。亚洲GDP前四名中的中国、韩国、印度聚到了一类,并且2001年-2016年是三个时期中中国距离众发达国家最近的时期,说明在数据科学成为独立学科后,亚洲的中国、韩国、印度等这些国家紧随发达国家之后大力发展数据科学。近几年随着大数据时代的到来,全球各国家对数据科学都极其敏感并且都正在迎合,大力发展数据科学。

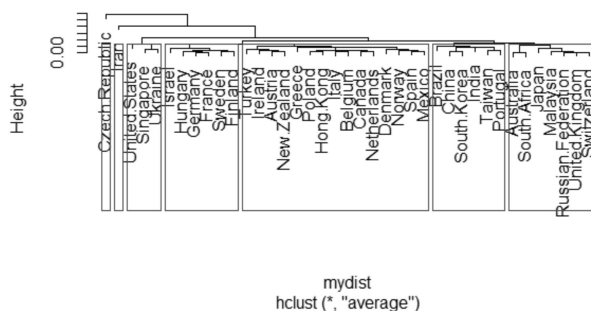


图4 2001年-2016年国家聚类图及其分类

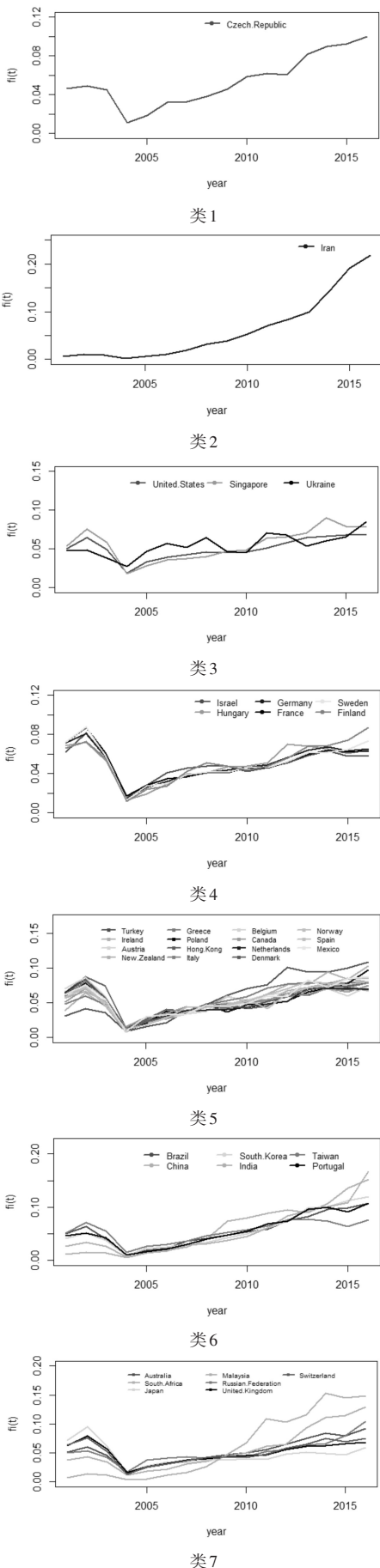


图5 2001年-2016年聚类各类别折线趋势图

比较三个时间段的聚类结果,每一个时间段,都会出现一个国家较多的类别,并且,该类别包括的国家大都在对应时期实力靠前。通过对三个时期的领头类别国家对比发现,加拿大、以色列、意大利、荷兰、比利时这些经济实力较强的国家始终都在该类别中,数据科学发展的三个不同时期,都存在一个当时实力领头的国家类别,随着时间发展,中国距离该类别越来越近,中国对数据科学的发展从起初的发展缓慢到最后的紧临领头国家类别,说明中国对一个新兴学科发展的实力所在。

### 5 论文来源机构分析

为了观察与数据科学相关机构的起伏变化,把自1980年至2016年中相邻两年的机构排名进行了比较。由于机构数量较多,选择了每年的发文量前30名进行了排名比较。比较两种排名之间的异同,可以根据逆序数<sup>[15]</sup>进行计算,  $T = |t_1| + |t_2| + \dots + |t_n| = \sum_{i=1}^n |t_i|$ , 其中  $t_i$  为一个机构在相邻两年的名次差,  $T$  为所有机构在两年中的名次变化总和。但是,运用逆序数会使得第二年未出现在前30名的机构难以计算。所以使用Jaccard距离来衡量相邻两年机构排名的变化程度:  $d_j = 1 - J$ , 其中Jaccard系数  $J(rank1, rank2)$ <sup>[16]</sup>通过给相邻两年的排名机构从前至后依次赋值30, 29, 28, ..., 0后计算

$$J(rank1, rank2) = \frac{\sum_{i=1}^{30} \min(x_i, y_i)}{\sum_{i=1}^{30} \max(x_i, y_i)}$$

得到。最后通过计算Jaccard距离来衡量相邻两年的机构变动程度。例如:当相邻两年的机构排名  $rank1, rank2$  相等时,  $J(rank1, rank2) = 1$ ,  $d(rank1, rank2) = 0$ ; 当相邻两年的机构排名始终不相等时,  $J(rank1, rank2) = 0$ ,  $d(rank1, rank2) = 1$ 。对相邻两年的机构排名变化程度进行量化得出折线图后,通过统计学中的滑动平均处理找到机构变化幅度大的几个年份和变化较大的机构。如图6所示,1981年-1982年,1986年-1987年,1994年-1995年,2006年-2007年,2011年-2012年的机构变化较大。针对这五年,将机构排名列举出来,在附录中可见放大后的机构比较图。通过比较,发现1996年第一次提出数据科学之前,机构排名的波动较大,造成机构排名波动的原因是大学发文量排名的起伏。在2001年数据科学成为一门独立学科之后,波动较大的两个阶段是2006-2007年和

2011-2012 年度,这两年机构间的波动幅度较小,并且机构变动中大学名次一般上升。比如:在 2006-2007 年的机构变动中,哥伦比亚大学上升 7 个名次,是这两年上升名次最多的机构;在 2011-2012 年的机构变动中,清华大学上升了 6 个名次,悉尼大学上升了 5 个名次。说明在高校在数据科学的文献计量机构中占了很大的比重,是发文献的主力军。从机构的排名变化中,中国科学院

发文量一直稳步增长,从 1995 年进入前 30,到 2003 年超过俄罗斯科学院一直居于第一。说明在研究数据科学的机构中,中国科学院的研究力量不容小觑。与此同时,通过图 7 机构变化曲线与幂函数的拟合 ( $R^2 = 0.5185$ ),可以看出,机构变化的波动幅度在逐渐降低。说明随着数据科学学科的发展与进一步改善,全球各机构也在不断完善与发展,机构的排名趋于平稳化。

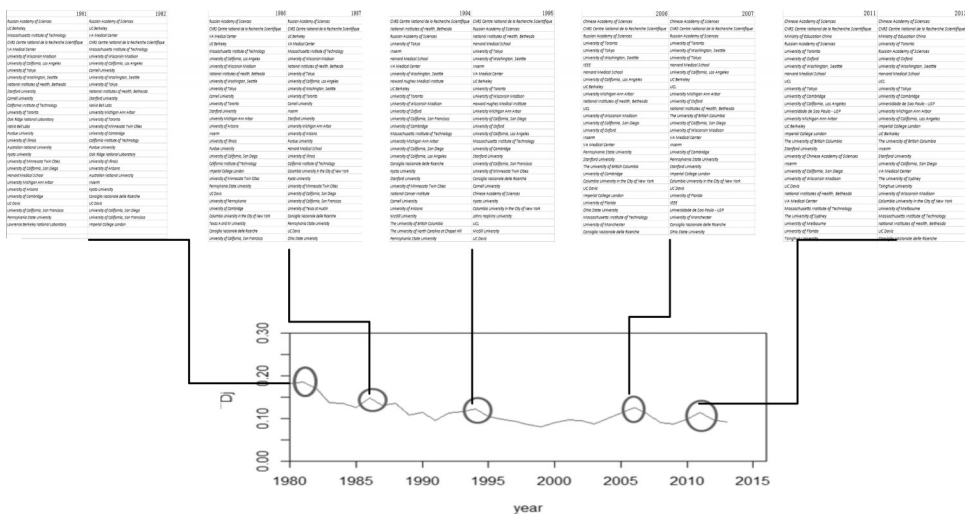


图 6 机构变化折线图

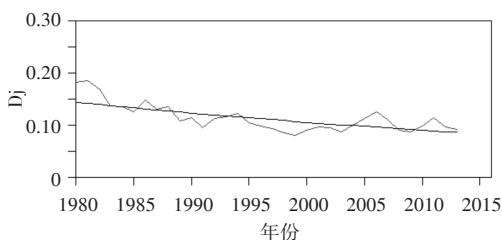


图 7 机构变化折线拟合图

## 6 结语

基于 SCOPUS 中与数据科学相关的论文数据,通过采用文献计量学的可视化分析,对国际上数据科学的论文总量、论文来源国家,以及论文来源机构按照不同的发展阶段进行了分析研究。首先对论文总量,通过指数拟合并建立回归模型对数据科学的论文发展进行预测,通过随机抽样验证预测正确性后发现数据科学论文在未来几年将持续指数型增长。显示了数据科学作为一个新兴学科的发展趋势。其次,通过不同时间段对论文来源国家聚类的方法,实现了对数据科学不同研究阶段的国家分析。通过三个阶段聚类以及类别走势分析,得出每个阶段数据科学领域的领头类别国家。并且随着时间发展,中国、韩国、印度等国家紧随发达国家之后,大力发展数据科学。比较三个不同阶段,

中国在数据科学领域从开始的边缘国家,起初的发展缓慢到数据科学成为独立学科后近几年紧随发达国家发展数据科学,可见中国在新兴学科的发展凸显出很强的后劲。最后,大数据时代的到来,在科学研究中以及实际应用中数据科学已经成为一个非常重要的分支,数据科学学科呈现出快速增长阶段。随着中国对数据科学的重视与发展,中国科学院也走到了机构中的前列。

## 参考文献

- [1] 杨瑞仙. 大数据研究的文献计量分析[J]. 情报科学, 2015(8):152-156.  
YANG Ruixian. Bibliometrical Analysis of the Study of Big Data [J]. Information Science, 2015(8):152-156.
- [2] 蒋德凤. ESI 学科动态评价与持续追踪分析模式研究——基于 Incites 学科评价角度[J]. 现代情报, 2018, 38(04):149-155.  
JIANG Defeng. Research on ESI Subject Dynamic Evaluation and Continuous Tracking Analysis Mode [J]. Modern Information, 2018, 38(04):149-155.
- [3] 刘婧. 文献作者分布规律研究——对近十五年来国内洛特卡定律、普赖斯定律研究成果综述[J]. 情报科学, 2004, 22(01):123-128.

- mobile social networks: Challenges survey and application in smart campus [C]// IEEE, International Conference on Wireless and Mobile Computing, NETWORKING and Communications. IEEE, 2016:1-8.
- [15] Zhu K, Li W, Fu X, et al. Data routing strategies in opportunistic mobile social networks: Taxonomy and open challenges [J]. *Computer Networks*, 2015, 93 (P1): 183-198.
- [16] Daly E M, Haahr M. Social network analysis for routing in disconnected delay-tolerant MANETs [C]// Acm International Symposium on Mobile Ad Hoc Networking & Computing. 2007:32-40.
- [17] Johari R, Gupta N, Aneja S. DSG-PC: Dynamic Social Grouping Based Routing for Non-uniform Buffer Capacities in DTN Supported with Periodic Carriers [M]. *Quality, Reliability, Security and Robustness in Heterogeneous Networks*. Springer Berlin Heidelberg, 2013: 1-15.
- [18] Johari R, Gupta N, Aneja S. CACBR: context aware community based routing for intermittently connected network [C]// ACM Symposium on PERFORMANCE Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks. ACM, 2013:137-140.
- [19] Hui P, Crowcroft J, Yoneki E. BUBBLE Rap: Social-Based Forwarding in Delay-Tolerant Networks [C]// ACM International Symposium on Mobile Ad Hoc NETWORKING and Computing. ACM, 2008:241-250.
- [20] Xia F, Ahmed A M, Yang L T, et al. Exploiting Social Relationship to Enable Efficient Replica Allocation in Ad-hoc Social Networks [J]. *IEEE Transactions on Parallel & Distributed Systems*, 2014, 25(12):3167-3176.
- [21] Li F, Wu J. LocalCom: a community-based epidemic forwarding scheme in disruption-tolerant networks [C]// Sensor, Mesh and Ad Hoc Communications and Networks, 2009. SECON '09. IEEE Communications Society Conference on. IEEE, 2009:1-9.
- [22] Li Y, Su G, Hui P, et al. Multiple mobile data offloading through delay tolerant networks [C]// IEEE, 2011: 43-48.
- [23] Samuel C Nelson, Bakht M, Kravets R, et al. Encounter: based routing in DTNs [J]. *ACM SIGMOBILE Mobile Computing and Communications Review*, 2009, 13 (1):56-59.

（上接第742页）

- LIU Jing. Study on Document Author's Distribution Regulation——A Review of 15 Years Research Articles on Lotka's & Price's Laws in China [J]. *Information Science*, 2004, 22(01): 123-128.
- [4] 邱均平, 杨瑞仙. 基于ESI的学科热门论文的计量研究——以临床医学学科为例 [G]. *情报科学*, 2010(1): 53-57.
- QIU Junping, YANG Ruixian. A Bibliometric Research on Hot Papers of Clinical Medicine Based on ESI [G]. *情报科学*, 2010(1): 53-57.
- [5] Gargiulo F, Caen A, Lambiotte R, Carletti T. The Classical Origin of Modern Mathematics [J]. *EPJ Data Science*, 2016, 5(1):26.
- [6] 赵蓉英, 魏明坤. 国际数据科学演进研究: 基于时间维度的分析 [J]. *图书情报知识*, 2017(4): 71-79.
- ZHAO Rongying, WEI Mingkun. International Evolution Research of Data Science: Based on the Time Dimension [J]. *Documentation, Information & Knowledge*, 2017(4): 71-79.
- [7] 李志毅. 计算机应用现状以及发展趋势 [J]. *电子技术与软件工程*, 2017(22): 155-155.
- LI Zhiyi. Current Status and Development Trend of Computer Applications [J]. *Electronic Technology & Software Engineering*. 2017(22): 155-155.
- [8] 赵庆澎. 计算机人工智能的发展现状与未来趋势 [J]. *电子技术与软件工程*, 2018(04): 254-254.
- ZHAO Qingpeng. Development Status and Future Trend of Computer Artificial Intelligence [J]. *Electronic Technology & Software Engineering*, 2018(04): 254-254.
- [9] 张海燕. 基于文献计量学的洛特卡定律研究综述 [J]. *中华医学图书情报杂志*, 2013, 22(08): 18-21.
- ZHANG Haiyan. Bibliometrics — based Lotka's law: A review [J]. *Chinese Journal of Medical Library and Information Science*, 2013, 22(08): 18-21.
- [10] 吴雨华. 普莱斯科学论文引用关联关系验证——以图书情报学17种核心期刊为例 [J]. *情报科学*, 2009(11): 1675-1678.
- WU Yuhua. Verification on Citation Relationship of Price's Scientific Paper—Taking 17 Core Journals of Information Sciences as a example [J]. *Information Science*, 2009(11): 1675-1678.
- [11] 杨京, 王效岳, 白如江, 等. 大数据背景下数据科学分析工具现状及发展趋势 [J]. *情报理论与实践*, 2015, 38(03): 134-137.
- YANG Jing, WANG Xiaoyue, BAI Zhujiang, et al. The status quo and development trend of data science analysis tools under the background of big data [J]. *Information Studies: Theory & Application*, 2015, 38 (03): 134-137.
- [13] Hayashi C, Yajima K, Bock H H, Ohsumi, N, Tanaka T, Baba Y (Eds.). *Data Science, Classification, and Related Methods* [C]// Proceedings of the Fifth Conference of the International Federation of Classification Societies (IFCS-96).
- [15] Fagin R, Kumar R, Sivakumar D. Comparing Top K-Lists [J]. *SIAM J Discrete Math.*, 2003, 17(1): 134-160.