基于CNKI数据库的我国各学科论文分布地图的绘制和分析

胡志刚[1] 陈华雄[2] 林歌歌[1]
（1. 大连理工大学科学学与科技管理研究所，辽宁大连 116000；2. 科技部科技评估中心，北京 100081）

摘要：为了分析和识别我国各省份的优势学科和研究领域，基于中国知网（CNKI）期刊数据库，通过统计2015年各省份所发表的论文量（按学科划分），绘制了我国的各学科论文分布地图。结果表明，数理学科主要分布在北方省份，生物化学主要分布在南方省份；重工业学科主要分布在北方省份，轻工业学科主要分布在南方省份；人文社科的研究主要集中在北京市。学科论文分布地图对于区域科技统筹发展具有一定的指导意义。

关键词：学科论文公布；分布地图；CNKI；学科；地图；可视化
中图分类号：G301 **文献标识码：**A **DOI：**10.3772/j.issn.1674-1544.2017.02.008

Analysis of Research Production Maps of Our Country Based on CNKI Database

HU Zhigang[1], CHEN Huaxiong[2], LIN Gege[1]
(1.Institute of Science Studies and S&T Management, Dalian University of Technology, Dalian 116000; 2. National Center for Science & Technology Evaluation, Beijing 100081)

Abstract: This research is conducted to identify the advantage research subjects of every regions in mainland China. In CNKI, we counted the amount of papers of each research subjects published by every Chinese provinces, cites and autonomous regions in 2015. For basic subjects such as mathematics and physics, its major producer is located in the north of China, while chemistry and biology is the priority of the southern provinces. For the engineering subjects, heavy industry related subjects are mainly produced by the northern provinces, light industry related subjects are mainly produced by the southern provinces. Beijing is the major producing region of Humanities and Social sciences. The generation of Chinese subject map has a certain guiding significance for the development of regional science and technology.

Keywords: subject paper distribution, subject maps, CNKI, subject, map, visualization

1 引言

在过去的 10 年里，中国知网（CNKI）收录的期刊论文量从 2006 年的 199 万篇上升到 2014 年的 272 万篇，增长了 36.7%。但由于不同省份的科研人员和资金投入规模悬殊，各地科技发展水平参差不齐。沿海地区的人才储备和水平较高，科研实力较强，而中西部地区由于受人才和

作者简介：胡志刚（1984—），男，大连理工大学科学学与科技管理研究所讲师，博士，研究方向：科学知识图谱、科学计量学；**陈华雄***（1975—），男，科技部科技评估中心副研究员，博士，研究方向：科技评估、科技管理研究；**林歌歌**（1992—），女，大连理工大学科学学与科技管理研究所硕士研究生，研究方向：科学计量学、科学学。

基金项目：中国博士后基金科学基金面上项目“基于文献计量的我国学科发展态势分析及评价”（2016M591435）。
收稿时间：2016 年 9 月 5 日。

资金的限制，科研水平相对比较落后。

Nature Publishing Group（NPG）发布的Nature Index 2015 China[1]显示：包括北京、上海、南京等在内的排名前10个城市发表的论文量占我国全部论文产量的73.2%。北京已经超过伦敦、东京、巴黎、纽约等城市，成为全球科研论文产量最高的研究中心[2]。根据Web of Science数据，2015年中国发表的SCI论文中约1/4的研究者来自北京。中国不同省市科研论文产出的巨大差异已经受到了国内外学者的广泛关注[3–5]。

全国各省市之间科技论文产出的差异不仅体现在数量上，而且体现在优势领域的构成方面。不同省市的研究侧重点可能大相径庭。以我国科研活动最活跃的北京和上海两个城市来说，北京有着全球最大的科研单位中国科学院的大部分研究所，在基础研究方面表现突出；而上海作为注重创业创新的国际化大都市，在生物技术等高技术领域实力更强。

从区域科技发展的角度，如何更好地评估和识别一个地区的优势学科和优势领域，紧紧围绕区域科技发展的高地和重点，发挥优势，突出特色，是区域科技发展过程中需要首先分析的关键问题。本文基于中国知网（CNKI）的数据，利用文献计量学和可视化的方法，统计了全国31个省、自治区、直辖市（不包括港澳台）在各个学科领域的发文量，绘制了各学科的分布地图。这对各区域了解自身科技发展中的学科结构和地缘优势，更好地进行科技投入和资源配置、找准区域科技发展的最优路径具有重要的价值和意义。

2　数据与方法

通过对科技文献进行统计和计量的方法测度一个地区的科研水平和表现，是科研评价中常用的一种方法。在中国科学院文献情报中心定期发布的《科学结构地图》系列丛书[6–8]中，就是利用了汤森路透科技集团的基本科学指标库（ESI）作为数据源，并借助文献计量学的理论和方法，应用多种可视化技术，揭示科学研究结构及其演变趋势。与《科学结构地图》相类似，很多学位论文也采用文献计量学的方法对我国的学科或领域的发展态势进行评估[9–11]。文献计量学方法减少了对领域专家的依赖，可以依靠方便易用的数据、方法和工具，对学科发展态势进行全面的监测和把脉。因此，这种方法越来越受到科学界的关注和认可。

本文采用的数据来源于中国知网（CNKI）的期刊全文数据库。CNKI期刊数据库是我国最大最全的期刊文献数据库，收录了近9000种学术期刊。这些期刊被划分为10个专辑，包括基础科学、工程科技、农业科技、医药卫生科技、哲学与人文科学、社会科学、信息科技、经济与管理科学等，各专辑又划分了若干学科和专业，共划分了168个学科和3000多个专业。这为我们分析各个学科的科技论文产出规模提供了一个全面而系统的学科分类体系和统计口径。

在中国知网中，借助其中的中国引文数据库进行数据统计。该数据库提供了文献导出、计量可视化、数据分析器等一系列分析功能。其中，数据分析器可以按照学科和地域分别进行统计。使用数据分析器下的地域分析器，可以对CNKI数据库中的数据按照地域进行统计分析，从而计量得到全国31个省、自治区、直辖市的学术产出和学术影响力以及各学科的地域分布状况。

具体操作如下：在数据分析器的检索页面上，地域选择列表中的“北京市”，出版时间选择“2006年”，期刊选择“核心期刊”，左侧选择“学科发文统计”，单击“分析”按钮并选择数据模式后得到的便是北京各学科2006年在核心期刊中发表的论文数量。按照此方法，下载2006—2015年全国31个省、自治区、直辖市在168个学科中的发文数据。

将下载得到的各区域和学科的发文量数据利用地理信息绘制工具进行数据可视化。笔者选取的是百度公司商业前端数据可视化团队所开发的ECHARTS平台。该平台是基于HTML 5 Canvas进行开发的，是一个纯Javascript图表库，提供直观、生动、可交互、可个性化定制的数据可视化图表。选取ECHARTS中适用的“地图”效果图，

下载代码并根据研究的需要进行修改，就可以得到2006—2015年各省市自治区的学科论文分布地图。

3 结果与分析

3.1 发文总量

统计2015年CNKI核心期刊数据库中各省、自治区、直辖市的发文总量，如表1所示。北京市以87718篇的论文产出量排名榜首，其论文产出量超过排名第二位江苏省约63.7%，上海论文产出量是30707篇，排名第三位。北京和上海是中国经济发达地区，而江苏省人均GDP在各省份中排名第一位，其省会城市南京市是长三角地区除上海之外第二大城市，拥有非常优良的经济、技术、教育和文化环境。根据教育部2016年公布的数据，江苏省拥有166所高等院校，位居全国首位，高校为江苏省科研产出贡献了很大的份额。

发文量最高的10个学科，如表2所示。可以看出，北京市的研究优势主要包括地质学、电力工业、环境科学与资源利用等领域；江苏的研究优势主要在中等教育、环境科学与资源利用、建筑科学与工程等领域；上海市的研究优势主要在建筑科学与工程、电力工业、环境科学与资源利用等领域；西藏自治区的研究优势主要在畜牧与动物医学、地质学和生物学等领域。

下面，从基础科学、工程技术、农林牧渔、医药卫生、人文社科和教育科学等六大门类中

表1 2015年我国各省市自治区的学术论文发文量

省份	论文量/篇	省份	论文量/篇	省份	论文量/篇	省份	论文量/篇
北京	87718	河南	23984	安徽	12474	云南	7735
江苏	53577	浙江	21206	黑龙江	12408	贵州	7185
上海	30707	辽宁	21015	江西	10411	内蒙古	6365
湖北	29131	河北	16582	福建	10363	海南	3028
广东	28546	湖南	16136	广西	9629	宁夏	2436
陕西	26195	重庆	14568	山西	8907	青海	1780
四川	25416	天津	14191	甘肃	8369	西藏	648
山东	25105	吉林	13186	新疆	7837		

数据来源：中国知网（CNKI）。

表2 2015年我国部分地区的主要高产学科

北京	江苏	上海	……	西藏
地质学（3785篇）	中等教育（2431篇）	建筑科学与工程（1530篇）	……	畜牧与动物医学（54篇）
电力工业（3419篇）	环境科学与资源利用（2044篇）	电力工业（1132篇）	……	地质学（43篇）
环境科学与资源利用（3419篇）	建筑科学与工程（1920篇）	环境科学与资源利用（1055篇）	……	生物学（39篇）
建筑科学与工程（2564篇）	高等教育（1880篇）	轻工业手工业（925篇）	……	农作物（35篇）
新闻与传媒（2398篇）	轻工业手工业（1880篇）	化学（876篇）	……	园艺（28篇）
金融（2344篇）	电力工业（1700篇）	金属学及金属工艺（833篇）	……	矿业工程（23篇）
计算机软件及计算机应用（2214篇）	计算机软件及计算机应用（1598篇）	计算机软件及计算机应用（810篇）	……	农业经济（19篇）
金属学及金属工艺（2045篇）	自动化技术（1377篇）	企业经济（775篇）	……	高等教育（19篇）
宏观经济管理与可持续发展（1986篇）	肿瘤学（1327篇）	金融（774篇）	……	计算机软件及计算机应用（19篇）
矿业工程（1974篇）	化学（1310篇）	公路与水路运输（760篇）	……	中药学（19篇）

数据来源：中国知网（CNKI）。

选取有代表性的学科，绘制并分析这些代表性学科的分布地图。这些代表性学科是从CNKI的168个学科中选取的，一般选取的是各学科门类下最基础的或者在区域分布上最具典型性的学科领域。

3.2　基础科学

以“数学”“物理”“化学”和“生物学”4个学科为案例，对基础学科的区域分布进行分析（图1）。这4个学科是传统的学科领域，也是其他众多理工科研究领域的学科基础。

“数学”领域是区域发展比较均衡的一个学科。在“数学”领域，除西藏、青海、宁夏外，各省份都有较好的表现。其中发文量相对较高的省市有：北京（613篇）、江苏（480篇）、陕西（446篇）、河南（388篇）、安徽（296篇）等，主要分布在长江以北、黄河以南之间的中部地区。“物理”领域的分布不均衡，发文量最多的有5个，即北京（933篇）、陕西（450篇）、江苏（446篇）、上海（381篇）和四川（377篇），远高于其余地区，其中又以北京市最为突出，它的发文量是排在第二位的陕西省的两倍多。

在“化学”领域，北京（1485篇）、江苏（1310篇）排在前两位并遥遥领先，上海（876篇）、广东（849篇）、辽宁（739篇）、浙江（716篇）、山东（707篇）等论文产出量次之，他们的发文量是北京市的一半左右。

在“生物”领域，排在前两位的同样是北京（1671篇）和江苏（1091篇），其次是广东（802篇）、山东（693篇）、上海（600篇）、湖北（523篇）、四川（509篇）等地区。

总的来说，北方省份（如陕西、河南）在“数学”和“物理”领域表现较好，而南方省份（如广东、上海）在“化学”和“生物”等应用性更强的领域表现较好，这反映了北方和南方不同的研究侧重和倾向。

3.3 工程技术

在工程技术门类中，分别从重工业、轻工业和电子计算机等主要技术方向中选取代表性学科

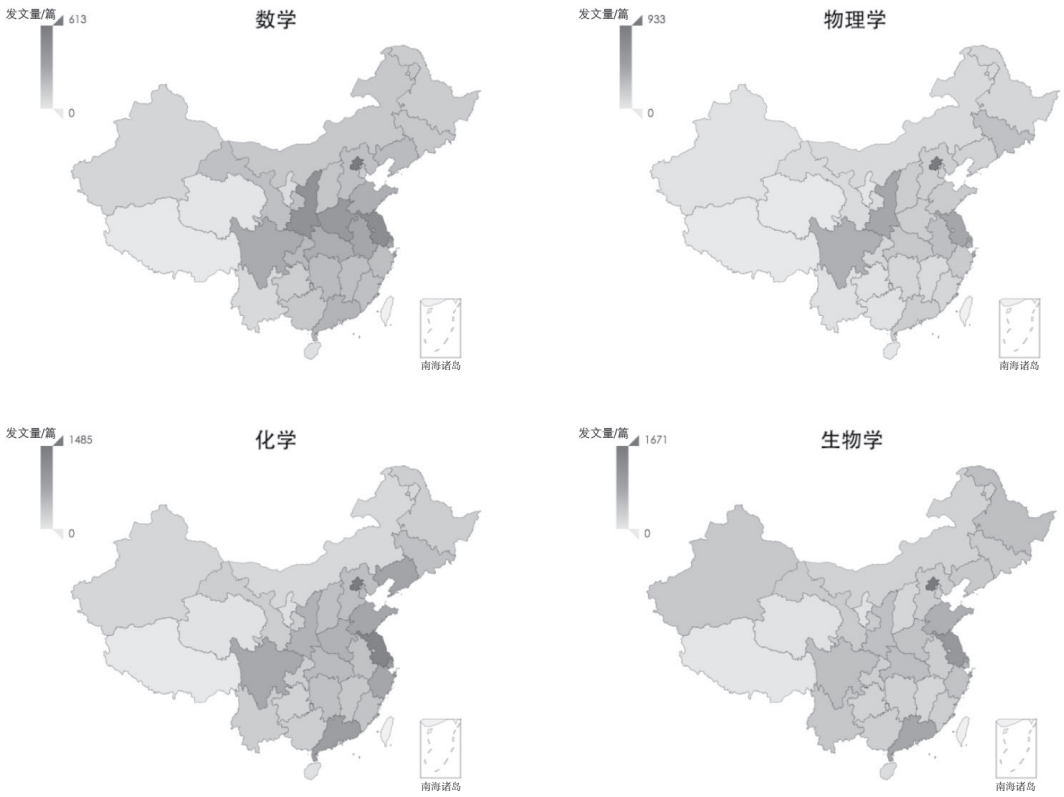

图1　基础科学领域代表性学科的论文分布地图

进行分析（图 2）。

在重工业领域，选取了“金属学及金属工艺”和“机械工业”两个学科。对这两个学科来说，北京市和江苏省仍然排在前两位，但是，陕西省和辽宁省相对于其他省份同样优势明显：在“金属学及金属工艺”中，辽宁省的发文量是 1317 篇，陕西省的发文量是 1031 篇；在“机械工业”领域，陕西省的发文量是 402 篇，辽宁省的发文量是 267 篇，明显高于其他地区尤其是南方省份。

在轻工业领域，与重工业领域的情况相反。南方省份具有一定的比较优势，发文量相对较高。在“有机化工”学科中，上海市和广东省分别以 718 篇和 608 篇的发文量排在仅次于北京市和江苏省的第三、四位；在“轻工业手工业”学科中，广东省以 1215 篇的发文量排在第三，浙江省以 1006 篇排在第四。显然，长三角和珠三角经济发达地区在轻工业领域具有较大的市场需求和较强的研究力量。

在电子计算机领域，其分布又不同于重工业和轻工业。凭借西安交通大学、西安电子科技大学、西北工业大学等多所工科院校的支撑，陕西

图 2　工程技术领域代表性学科的论文分布地图

省在这一领域表现突出，牢牢锁定除北京市和江苏省外的第三位置。此外，上海、湖北、四川、河南等省市在该领域也有不错的表现。

由此可以看出，工程技术领域的分布受经济因素的影响：一是经济发展水平，工业发达地区的工程技术研究水平总体高于其他欠发达地区；二是经济结构特点，北方省份更侧重重工业，南方更侧重轻工业。此外，在我国高校中存在着大量的工科类院校，这些院校的分布在很大程度上决定了今天我国主要工程技术学科的分布情况。

3.4　农林牧渔

相对于基础学科和工程技术，农业学科主要涉及农作物、畜牧与动物医学、林业、水产和渔业，其分布总体比较均衡（图 3）。由于欠发达省份对农业的依赖程度和重视程度更高，拉平了与发达省份之间在农业领域发文量的差距。

从“农作物”研究的情况来看，虽然北京（978 篇）、江苏（787 篇）的发文量仍然居于前两位，但是河南（624 篇）、四川（496 篇）、云南（445 篇）、甘肃（432 篇）、湖北（420 篇）、新疆（414 篇）、黑龙江（412 篇）等农业大省的发文量也都在 400 篇以上，尤其是云南、甘肃、新疆、黑龙江在总体科研产出不高的情况下，尤其显示出其在农业领域的突出地位。

“畜牧与动物医学”领域的情况与农业相类似，各地之间的发文量差距不大。除了北京、江苏外，黑龙江（566 篇）、山东（557 篇）、新疆（512 篇）、吉林（500 篇）等地区的发文量也超过了 500 篇，排名比较靠前。可以看出，北方省份在“畜牧与动物医学”领域的发文量明显高于南方各省份。

在“林业”领域，北京市以 1055 篇的发文量一枝独秀，比排在第二位的黑龙江（387 篇）高了近两倍。林业学科还是一个依赖于地区资源

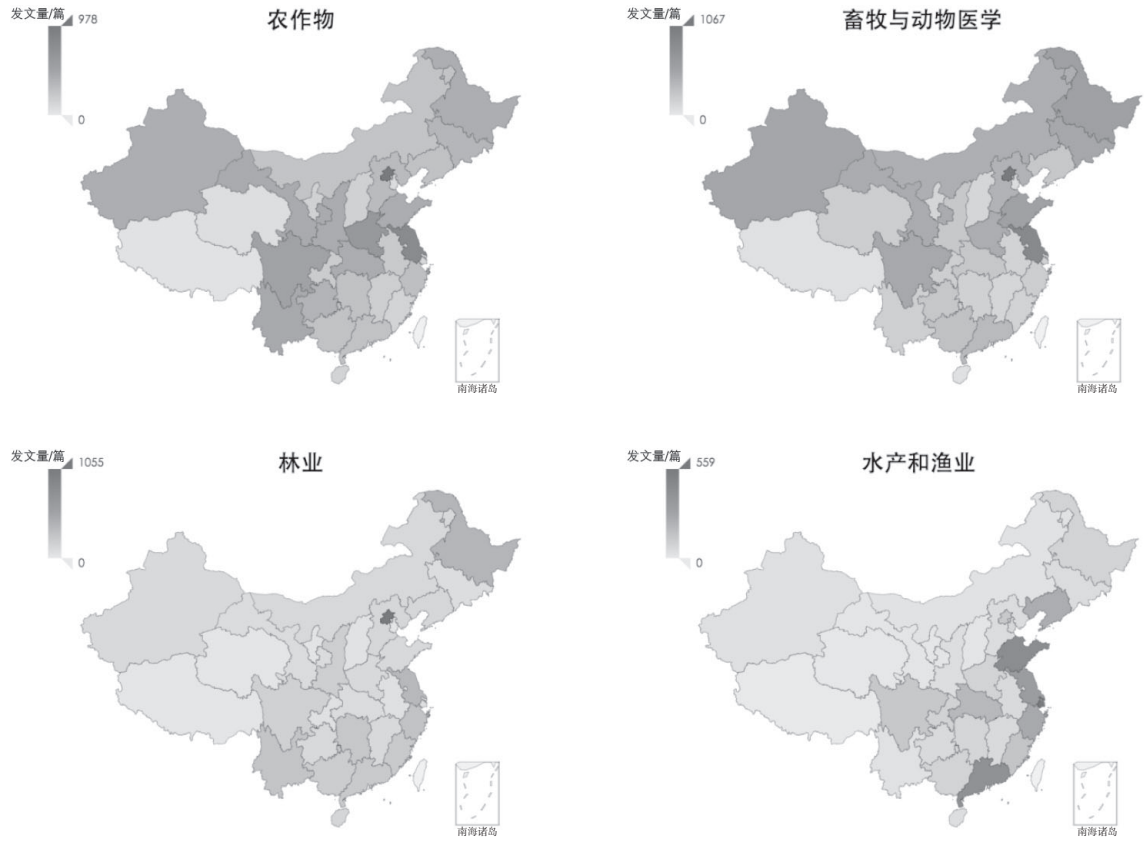

图 3　农林牧渔领域代表性学科的论文分布地图

禀赋的学科。例如：发文量排在第二的黑龙江省是林业大省，森林面积覆盖率全国第一，曾长期作为我国最大的木材生产和加工基地。

与“林业”相似的是“水产和渔业”学科。“水产和渔业”同样依赖于地理资源。显然，沿海地区在“水产和渔业”领域具有天然优势，是发文量最高的地区，如上海（559篇）、山东（432篇）、广东（399篇）、江苏（324篇）、辽宁（238篇）和浙江（254篇）都属于沿海地区。此外，位于长江流域的内陆省份湖北、四川也有较好的表现，而发文总量居榜首的北京市则由于水资源缺乏表现相对较弱。

农林牧渔领域的论文分布情况在一定程度上反映了我国农林牧渔主产区的分布情况。由于北方省份拥有我国最大的平原、牧场和林场，因此在这些领域相对于南方省份具有一定的比较优势，而在水产和渔业领域，沿海地区则比内陆地区具有天然的资源禀赋条件，科研表现自然也明显领先。

3.5 医药卫生

在医药卫生门类下，选取了“临床医学”和“肿瘤学”领域作为案例进行分析（图4）。对于“临床医学”领域来说，发文量最多的是江苏（973篇）。江苏省的医学院有南京医科大学、中国药科大学、南京中医药大学、徐州医科大学等，且有众多实力强劲的地方医院，因此，以多取胜，发文量排在了全国各省市之首，超过了医疗水平更高的广东（825篇）和北京（783篇）。“肿瘤学”的情况与“临床医学”相同，江苏省凭借1327篇的发文量排在首位，排在第二位和第三位的同样是广东（1108篇）和北京（867篇）。

由此可以看出，北京、江苏、广东等发达省份构成了我国医药领域研究的第一集团，发文量明显高于其他省份。而辽宁、河北、山东、河南、湖北、四川等人口大省以及上海市的发文量也比较高，处于医药类论文的第二集团。

3.6 人文社科

在人文社会科学领域，选择了“中国文学”“哲学”“金融”和“行政法及国家行政管理”4个代表性学科（图5）。北京市的发文量遥遥领先于其他省市，显示出北京地区在人文社科领域的绝对领先地位。

在“中国文学”领域，北京市的发文量为980篇，比排在第二位的江苏（583篇）多了近70%；在“哲学”领域，北京市的发文量是587篇，比排在第二位的上海（273篇）多了一倍还多。在“行政法及国家行政管理”领域，北京市的发文量（1212篇）比排在第二位的江苏省（453篇）高出168%。在“金融”领域，北京市的发文量（2344篇）是排在第二位的上海（774篇）的三倍。

显然，北京市在人文社科领域具有垄断性优势。一方面是因为北京市有很多以文科见长的高校，如北京大学、中国人民大学、北京师范大

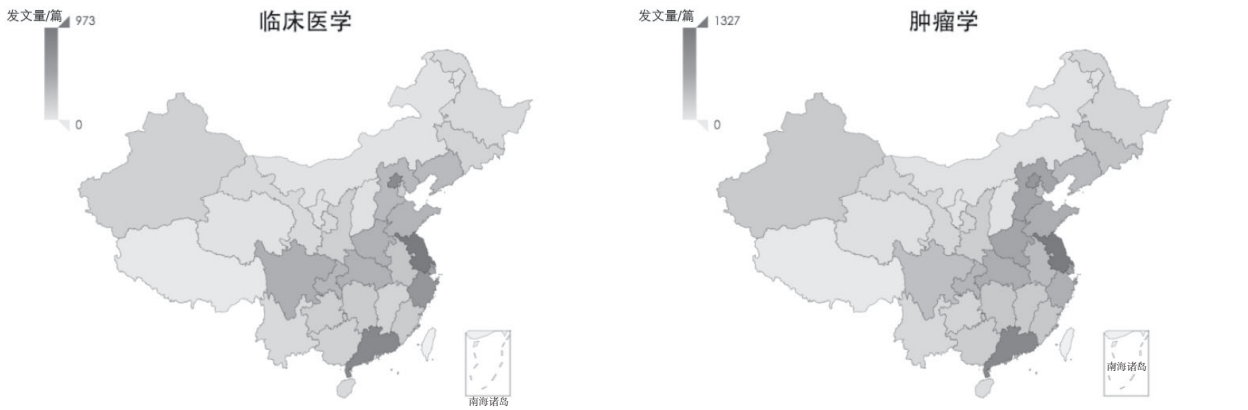

图4 医药卫生领域代表性学科的论文分布地图

学、北京语言大学、北京外国语大学、中国政法大学、对外经济贸易大学等，还有中国社会科学院、国家行政学院等国家级研究机构；另一方面是因为北京市作为我国的首都，是我国的政治中心、文化中心和经济中心，在政策资源和人才储备方面具有得天独厚的优势。

3.7　教育科学

选取“高等教育”和“职业教育”两个学科的发文量对教育学科论文的分布情况进行观察分析（图 6）。

在“高等教育”领域，江苏省以 1880 篇的发文量排在首位，略高于北京的 1756 篇。此外，湖北（860 篇）、上海（695 篇）、浙江（694 篇）、广东（684 篇）、河南（668 篇）、山东（651 篇）和四川（604 篇）的发文量也都在 600 篇以上。可以看出，“高等教育”的发文量与我国高等教

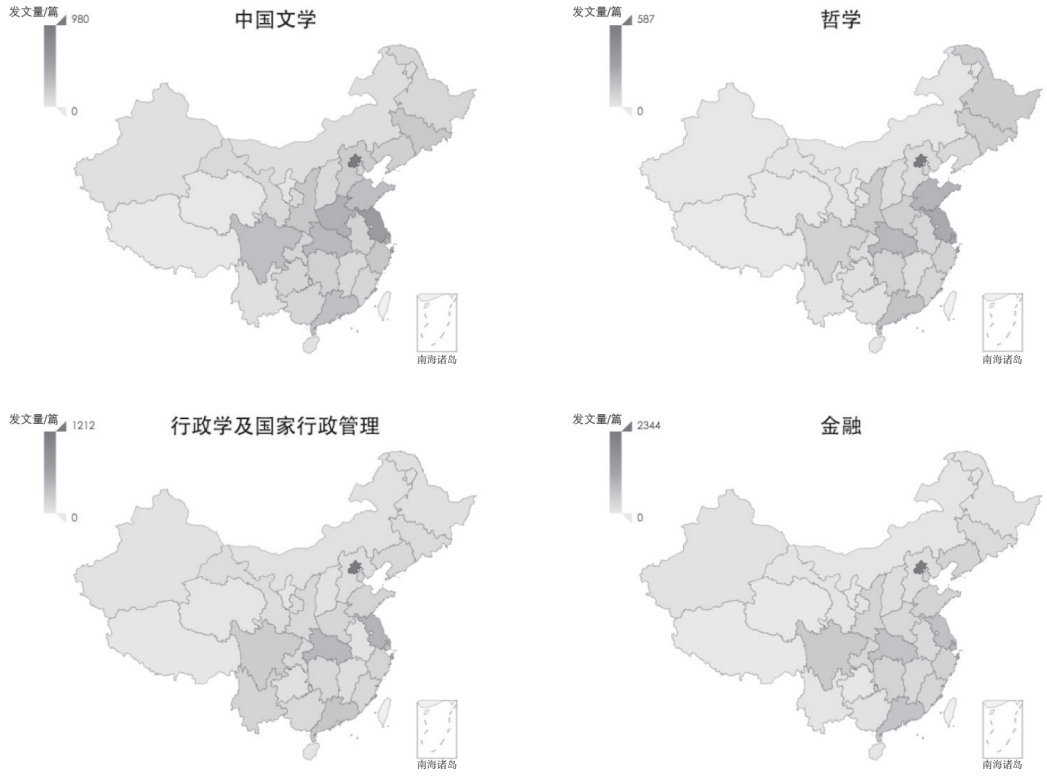

图 5　人文社科领域代表性学科的论文分布地图

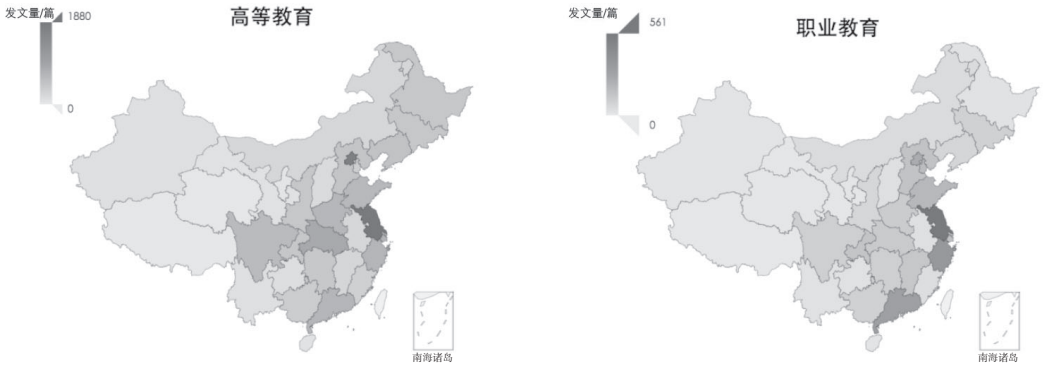

图 6　教育学领域代表性学科的论文分布地图

育的分布情况比较一致，高校较多的省市，如江苏（77所）、湖北（68所）、山东（67所）、北京（66所）、辽宁（65所）和广东（62所）等，在高等教育研究领域中表现均较为突出。

“职业教育”的情况与“高等教育”相类似，发表相关论文较多的省份，如江苏（561篇）、浙江（357篇）、广东（304篇）等，其职业教育也比较发达。这3个省份所拥有的专科院校数量分别位居全国第一位（89所）、第十三位（48所）和第二位（85所）。

4 讨论与结论

利用可视化的方法，按照CNKI中核心期刊论文的发文量，绘制理、工、农、医和人文社科类学科期刊论文的分布地图，展现了我国各学科地域分布的主要特点：一是科研分布总体不平衡，西部地区和南方地区发文产出较少，华北、华东地区学术产出较高，北京和江苏两地的发文量更是远高于其他省市自治区；二是北方与重工业相关的学科比较强，南方与轻工业相关的学科比较强，理科即基础学科的分布比较平均；三是农林牧渔学科的产出与我国第一产业的分布相类似，受地区地理位置和资源禀赋的影响较大；四是医药学科从北向南形成了北京、江苏和广东3个学术高地，其他人口大省亦有不俗的表现；五是北京市在人文社科类学科中处于绝对的领先地位；六是教育学科的学术论文产出数量与该地区的高校数量密切相关。

各学科论文分布地图反映了科学、技术与经济社会发展之间的关系。正如默顿在其《17世纪英格兰的科学、技术与社会》[12]一书中所分析的那样，17世纪英格兰在纺织业、采矿业和交通运输业的发展对英格兰在动力学、热力学等领域的科技进步具有重要的推动作用。同样，我国各省市的经济构成、资源禀赋、地理环境以及军事政治因素均对学科发文量具有多元复杂的影响。

学科论文发文量的差异性反映了各省市自治区科研水平分布的不平衡性，既需要协调发展，统筹推进，又需要因地制宜、找准优势。2015年11月5日，国务院发布了《统筹推进世界一流大学和一流学科建设总体方案》。该方案提出要以学科为基础，引导和支持高校优化学科结构，凝炼学科发展方向，突出学科建设重点。各省市自治区的科技发展也是一样，要根据自身所拥有的资源和条件，充分挖掘和利用自身资源，找准和抓住自身优势，提高优势领域的科技发展水平，为服务当地经济社会发展做出更大的贡献。

参考文献

[1] MCGHEE K, PHILLIPS N. Nature index 2015 China[J]. Nature, 2015, 528(7582 Supplement): 165–195.

[2] ANDERSSON D E, ANDERSSON A E, MELLANDER C. Handbook of creative cities[M]. Cheltenham: Edward Elgar Publishing, 2011.

[3] ANDERSSON D E, GUNESSEE S, MATTHIESSEN C W, et al. The geography of Chinese science[J]. Environment and Planning A, 2014, 46(12): 2950–2971.

[4] BORNMANN L, DE MOYA ANEGÓN F, LEYDESDORFF L. The new excellence indicator in the world report of the SCImago institutions rankings 2011[J]. Journal of Informetrics, 2012, 6(2): 333−335.

[5] LEYDESDORFF L, PERSSON O. Mapping the geography of science: distribution patterns and networks of relations among cities[J]. Journal of the American Society for Information Science and Technology, 2009, 61(8): 1622−1634.

[6] 潘教峰, 张晓林, 王小梅, 等. 科学结构地图2009[M]. 北京: 科学出版社, 2010.

[7] 潘教峰, 张晓林, 王小梅, 等. 科学结构地图2012[M]. 北京: 科学出版社, 2013.

[8] 王小梅, 韩涛, 王俊, 等. 科学结构地图2015[M]. 北京: 科学出版社, 2015.

[9] 于江. 学科领域发展状况的可视化分析[D]. 大连: 大连理工大学, 2009.

[10] 张静. 学科交叉分析方法与应用研究[D]. 大连: 大连理工大学, 2010.

[11] 薛萍. 基于社会网络的研究型高校优势学科方向发展预测研究[D]. 杭州: 浙江大学, 2014.

[12] 默顿. 17世纪英格兰的科学, 技术与社会[M]. 北京: 商务印书馆, 2009.