

江苏省大数据产业空间分布特征分析

朱建军¹

(无锡商业职业技术学院 商学院, 江苏 无锡 214153)

【摘要】: 数据可以促进传统产业服务个性化, 推动企业决策智能化, 深入研究大数据产业的空间特征对区域内产业空间结构调整与优化, 促进区域数字经济发展有重大意义。运用平均最近邻分析与核密度分析方法研究江苏省市两级大数据产业的空间分布特征。泰州市平均最近邻比率 $ANN=1.002307$, P 值=0.9878, 接受其空间统计意义的随机集聚; 核密度分析下的南京属于高核密度地区, 苏州、盐城次之。江苏省大数据产业集聚现象尚未形成, 但有加速集聚趋势; 长江流域大数据产业连绵区与苏北大数产业带初现。未来要积极探索江苏省大数据产业发展模式, 实现大数据产业的科学布局, 促进大数据产业地区间的融合与协同发展。

【关键词】: 江苏 大数据产业 空间分布

【中图分类号】: F49 **【文献标志码】:** A **【文章编号】:** 1671-4806(2019)06-0001-05

伴随着信息技术的高速发展, 数据成为促进现代经济社会发展的关键因素, 成为国家基础性战略资源。全球数据呈现爆发增长、海量集聚的特点, 世界各国都把推进经济数字化作为实现创新发展的重要动能。大数据作为数字化转型中的核心, 相关技术和应用助推了数字中国战略的加速落地。根据国家网信办的数据, 2018 年中国数字经济规模达 31.3 亿元, 同比增长 15.1%, 占 GDP 的比重达 34.8%。受宏观政策、技术升级和应用场景等因素影响, 2018 年我国大数据产业规模为 4384.5 亿元, 预计 2021 年将达 8070.6 亿元。2016—2021 年, 大数据产业规模增长 5230 亿元, 5 年复合增长率达 23.2%^[1]。但目前各地大数据产业规划布局表现出比较明显的同质性与不平衡性, 本文聚焦江苏省大数据产业的空间布局进行分析具有现实意义。

一、文献回顾

目前, 国内关于大数据产业的研究还处于初步探索阶段, 主要集中在三个方面: (1) 基于大数据产业类型、特点、现状发展大数据产业的策略研究。迪莉娅^[1]、朝乐门^[2]、沈俊鑫和陈颖谦^[3]等人分析了大数据产业的特点、分类、现状与发展策略。(2) 借鉴发达国家和我国贵阳市发展大数据产业的经验, 发展我国大数据产业的研究。汪晓文等人通过波特钻石模型对日本、美国和中国大数据产业进行竞争优势比较研究, 认为我国的大数据产业尽管处于发展的初级阶段, 但形成了中国的特色和特征^[4]。谢卫东等人从大数据产业发展顶层设计、要素供给、数据隐私与安全保护以及大数据标准体系等四个方面对国内外大数据产业政策进行比较, 指出从数据开放和个人隐私保护政策完善、数据流通体系建立、人才培养和大数据关键技术投入四个方面促进我国大数据产业有效、健康发展^[5]。(3) 大数据产业统计核算研究。周曙东按《国民经济行业分类标准》将大数据产业的统计范畴分为核心活动和扩展活动两大类: 核心活动定义为主产品为大数据产品或其业务支撑大数据的产出、传输、组织管理及价值实现; 扩展活动指主业中与大数据产生、传输、组织管理及价值实现有关但又未列入核心产业部门的单位利用大数据在各领域的实际应用^[6]。周曙东利用大数据投入产出表对“非纯粹”大数据产业部门规模进行测度研究^[7]。

区域特色创新发展促进大数据产业快速聚集, 但国内学者对大数据产业聚集测度、聚集模式、聚集效应等尚未深入研究,

¹基金项目: 江苏高校哲学社会科学项目“大数据产业空间集聚与区域技术创新效率研究”(2018SJA0835)

作者简介: 朱建军(1976—), 男, 江苏如东人, 讲师, 硕士, 研究方向为经济与金融理论、计量经济。

事实上详细研究我国大数据产业的空间特征，对大数据产业空间布局与规划，促进大数据产业健康发展，促进区域经济发展具有重要的意义。本文运用平均最近邻分析和核密度分析方法探索江苏省大数据产业发展模式，这样的研究，有助于促进大数据产业区间的融合与协作发展。

二、产业空间集聚测度常用方法

（一）平均最近邻比率（ANN）

平均最近邻比率是每个要素的实际观测平均距离与基于随机分布假设的期望平均距离之商，其计算公式为：

$$ANN = \frac{S_0}{S_E} = \frac{2 \sum_{i=1}^N D_i}{\sqrt{NA}} \quad (1)$$

式（1）中，N 为区域要素数量，A 为所有要素包络线面积， D_i 为要素 i 与其最近邻要素的距离。通过 $ANN > 1$ 或 $ANN < 1$ 可判断我国各区域大数据产业是否存在集聚现象，并且判断区域集聚程度大小。

（二）核密度估计法

核密度估计法是一种非参数估计方法，借助一个移动的单元格，对给定区域的点或线格局的密度进行估计，用测量来的点或线生成连线表面，找出哪些地方点或线比较集中，计算整个区域的数据聚集状况。密度函数的估计方程为：

$$f_{h(x)} = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (2)$$

式（2）中， $K(*)$ 为核密度形式，h 为窗宽， $x - x_i$ 表示估计点 x 到已知 x_i 点的距离。

（三）全域空间自相关和局域空间自相关

全局 Moran' sI 指数是最早应用于全局聚类检验的方法，其可以检验整个研究整个区域中邻近地区间是相似、相异（空间相关、负相关），还是相关独立的。全局 Moran' sI 指数的计算公式如下：

$$\begin{aligned}
\text{Moran's I} &= \frac{n \sum_{i=1}^n \sum_{j=1}^j w_{ij} (x_i - \bar{x})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^n \sum_{j \neq 1}^n w_{ij} (x_i - \bar{x}) (x_j - \bar{x})}{S^2 \sum_{i=1}^n \sum_{j=1}^n w_{ij}} \quad (3)
\end{aligned}$$

式(3)中, n 是研究区域内地区总数; w_{ij} 是空间权重矩阵(设区域 i 和区域 j 相邻时, $w_{ij}=1$; 区域 i 和区域 j 不相邻时, $w_{ij}=0$); x_i 和 x_j 分别是区域 i 和区域 j 的属性; S^2 是属性的方差。

局部 Moran' sI 可以用来检验局部地区是否存在相似或相异的观察值集聚在一起, 区域 i 的局部 Moran' sI 用来度量区域 i 和它相邻区域之间的关联程度, 计算公式为:

$$\text{局部 Moran's I} = \frac{(x_i - \bar{x})}{S^2} \sum_{j \neq i} w_{ij} (x_j - \bar{x}) \quad (4)$$

式(4)值大于零, 表示一个高值被一个高值包围或一个低值被一个低值所包围; 值小于零, 表示一个高值被低值所包围或一个低值被高值所包围。

(四) 多距离空间聚类分析

多距离空间聚类用来确定要素(或与要素相关联的值)是否在某一距离范围内显著聚类或离散, 使用 Ripley' sK 函数公式为:

$$L(d) = \sqrt{\frac{A \sum_{i=1}^N \sum_{j=1, j \neq i}^N k(i, j)}{\pi N(N-1)}} \quad (5)$$

式(5)中 A 表示区域, N 表示点数, d 表示距离, $k(i, j)$ 表示权重, 当 i 和 j 之间的距离小于或等于 d 时, 权重等于 1, 当 i 和 j 之间的距离大于 d 时, 权重为 0。

三、实证研究

（一）样本选取

根据上海市新兴行业分类指导目录（2017 版）中的新兴行业大数据服务行业分类与代码（6540 数据处理和存储服务），在国家企业信用信息公示系统（江苏）中以“数据”作为关键词，以地级市和成立 1 年以内（剔除成立时间为 2019 年）和成立 1~5 年为查询条件，剔除列为经营异常和注销中的企业，逐一核对企业经营范围，最后遴选 534 家大数据产业企业。利用百度地图坐标拾取器，按企业登记地址收集样本企业的经度和纬度坐标，同时采集各企业的成立时间、注册资本和股东类型等要素。

（二）江苏省大数据产业空间分布特征分析

1. 江苏省大数据产业整体空间分布

Arcgis 软件中的要素有三种类型，分别为点要素、线要素和面要素。把每个大数据产业企业抽象为点要素并且落实在江苏省地图上，得到截至 2018 年的江苏省大数据产业空间分布结果，即江苏省大数据产业企业主要分布在苏南、苏中地区。2014—2018 年江苏省大数据产业企业注册资本由 25230 万元快速上升到 2018 年的 409819 万元，年均增长率为 253%，大数据产业得到快速发展（见图 1）。截至 2018 年底，江苏省各地级市大数据产业累计注册资本金额排前六位的分别是苏州 260079 万元、南京 231837 万元、无锡 172710 万元、徐州 137710 万元、盐城 82601 万元、南通 75407 万元（见图 2）。

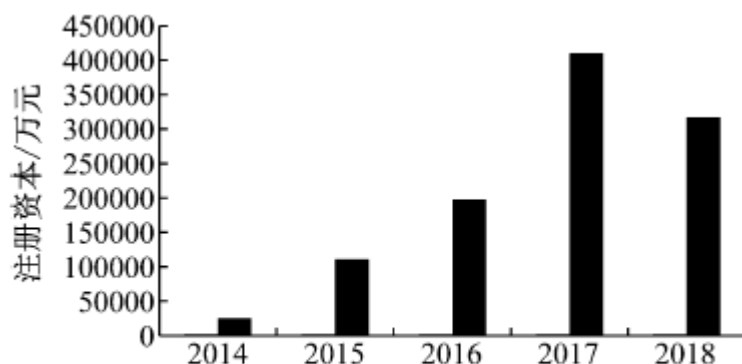


图 1 江苏省 2014—2018 年大数据产业投资额

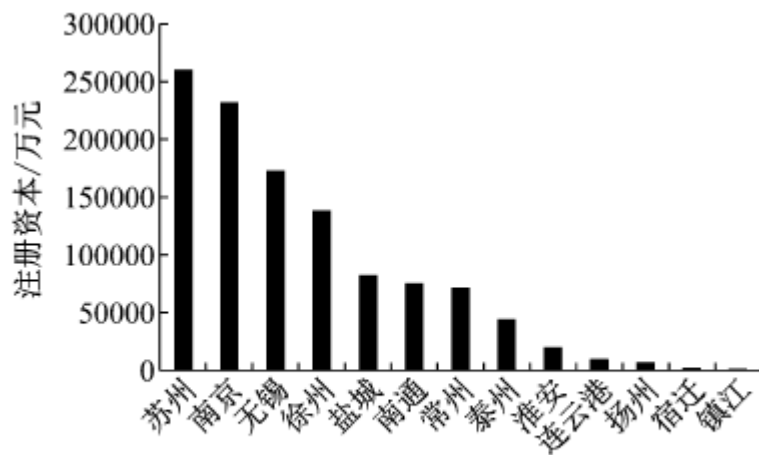


图 2 截至 2018 年的江苏市域大数据产业累计投资

2. 江苏省大数据产业空间聚集现象探索

(1) 泰州市健康大数据产业集聚初现。利用 Arcgis10.5 的空间分析工具对最邻近距离进行测算，得出实际最邻近距离均值、理论邻近距离均值及两者之比 R 、 Z 值和 p 值，运算结果生成报表。结果显示，江苏省市两级大数据产业的最邻近距离比率均大于 1，说明江苏省市两级的大数据产业尚未出现最邻近距离测算下的集聚现象。其中江苏省范围内的最邻近距离比率为 1.717498，13 个市的最邻近距离比率范围为 1.002307 至 2.520398，比率值最大的是连云港市，最小的是泰州市。从 Z 值来看，除泰州市外其他 12 个市的 Z 值均大于 2.58，进一步说明这 12 个市大数据产业空间分布呈分散状态。

根据泰州市最邻近距离判断空间集聚识别指标：最邻近距离比率为 1.002307， Z 值为 0.015291， P 值为 0.987800，可以认为泰州市大数据产业已经初步出现集聚现象，但在空间分布属于空间随机分布。与省内其他兄弟市相比，泰州市于 2018 年初出台了《泰州市大数据发展规划（2018—2020 年）》，在泰州数据产业园基础上打造华东地区最具特色、最具竞争力的健康医疗大数据产业。2018 年 4 月泰州医药高新区发布了《2018 年大健康产业集聚发展工作要点》，进一步突出大力促进资源要素集聚，促进泰州健康医疗大数据产业集聚。

(2) 江苏省大数据产业空间分布特征。利用 Arcgis10.5 软件中核密度分析工具生成江苏省大数据产业空间分布核密度估计值，按江苏省大数据产业核密度估计值九级分类，各地市的核密度存在差异。具体而言，南京是全省核密度值最高的地区，以核密度九级形成江苏省为首的大数据产业聚集地区，并辐射镇江句容地区。苏州核密度为八级，无锡核密度为六级，常州核密度为五级，形成大数据产业核心带。盐城以核密度七级、南通与徐州以核密度五级分布。进一步可以概括江苏省大数据产业空间分布总体特征为：南京-镇江地区、苏锡常地区、扬州-泰州地区、南通地区形成江苏长江流域大数据产业连片区；苏北地区以盐城、徐州为两核，其他地区零星分布。

四、江苏省大数据产业空间集聚模式分析

一是市场创造型产业集聚模式。从大数据产业专业化角度分析，大数据公司和产业活动在城市空间上必定集中在一定的区位，不同专业类型的活动趋向于在不同的地区集聚，城市空间的异质性在很大程度上与城市主导产业专业化联系十分紧密，大数据产业空间集聚属于城市地区专业化的一种现象，泰州大数据产业集聚就属于市场创造模式。在泰州医药高新区首先出现了医药产业的专业化市场，为健康医疗大数据产业集聚的形成创造了重要的市场交易条件和信息条件。

二是经济关联型产业集聚模式。范围经济由于市场的不确定性和技术的变化开始削弱，由此引起水平的和垂直的生产分离或生产的外部化，从而有更大的能力满足不同的需求和对市场力更大的适应性。哪里产生地理上敏感的交易成本多样化联系，哪里的外部化必定与集聚相联系。产业体系的灵活采用与专业化产业集聚体中生产空间的集中相联系，公司紧密地集聚以保证面对面的交流。这种集聚的目的在于加强地区内企业之间的经济联系，为企业发展创造更有利的外部条件。南京-镇江和苏锡常大数据产业集聚模式属于经济关联型。

三是资本转移型集聚模式。这种模式一般是发生在有产业转移或新兴产业产出的背景下，当一个规模较大的企业出于接近市场或节约经营成本或获取政策优惠的考虑，在生产区位上做出重新选择，并投资于一个新的地区的时候，有可能引发同类企业及相关企业朝这个地区汇集。这种产业集聚的形成，主要缘于资本迁移和资本流动而形成的产业集聚现象。盐城大数据产业园是国家级园区，有多种政策优惠，吸引了众多大数据产业企业的汇集。

五、结论与建议

（一）结论

江苏大数据产业空间聚集现象尚未形成，但泰州在最邻近距离分析测算下接近空间统计意义上的聚集。大数据产业空间分布呈现长江流域连绵区，苏北地区呈现以盐城和徐州两核、其他零星分布的空间结构。

（二）建议

1. 探索江苏省大数据产业“一区一带”的空间格局发展模式

江苏省大数据产业长江流域连绵区初步形成，要积极探索南京和苏州的龙头作用，加强区内各大数据产业聚集核间的关联与协调。发挥盐城和徐州两聚集核的辐射效应，聚焦打造盐都区—淮阴区—宿豫区—铜山区苏北大数据产业发展带。

2. 探索“核心驱动、分业集聚”的泰州大数据产业发展布局模式

泰州大数据产业聚集现象领先全省，得益于“核心驱动，分业集聚”的“泰州模式”。核心驱动即打造大数据产业核心区，发展大数据服务业；分业集聚即以医药为高新区特色产业，发展健康医疗大数据应用。“泰州模式”很好地实现了健康医疗产业要素向医药高新产业集聚，实现了分布合理、功能错位、特色发展的大数据集聚发展态势，其他地级市必须认真借鉴“泰州模式”，形成自己的产业发展布局模式。

3. 推动市际协同创新的大数据产业空间关联

在国家大数据产业布局的大框架下，各地级市要结合自身所处的产业链位置，明确其功能定位，结合主导产业发展切入点，探索大数据驱动发展模式。在此基础上，各地方政府要打破行政区划壁垒，整合优势资源促进区域合作，形成大数据产业协调发展的格局。

参考文献:

[1]迪莉娅.我国大数据产业发展研究[J].科技进步与对策,2014(4):56-60.

[2]朝乐门.我国大数据产业的特征分析与政策建议[J].情报理论与实践,2016(10):5-10.

[3]沈俊鑫,陈颖谦.面向欠发达地区大数据产业发展能力分析的网络化方法研究[J].通信学报,2017(12):153-159.

[4]汪晓文,曲思宇,张云晟.中、日、美大数据产业的竞争优势比较与启示[J].图书与情报,2016(3):67-74.

[5]谢卫红,樊炳东,董策.国内外大数据产业发展比较分析[J].现代情报,2018(9):113-121.

[6]周曙东.我国大数据产业统计范畴的分析与探讨[J].调研世界,2017(12):49-52.

[7]周曙东.大数据产业投入产出表构建及经济贡献测试方法研究[J].调研世界,2019(3):7-10.

注释:

1 数据来源于大数据产业生态联盟联合赛迪顾问共同完成的《2019 中国大数据产业发展白皮书》。但由于大数据产业统计范围尚未统一,不同研究机构统计的大数据产业规模存在差异。中国信息通信研究院发布的 2018 年我国大数据产业规模突破 6000 亿元。