

全国数字经济发展综合评价

——基于省级面板数据的分析

蔺泽华, 刘海东, 李 姝

天津商业大学理学院, 天津

收稿日期: 2023年9月23日; 录用日期: 2023年10月14日; 发布日期: 2023年10月30日

摘 要

本研究探讨了中国数字经济的发展, 并构建了一个综合评价体系, 基于2013~2020年的省级面板数据。通过运用主成分分析、K-means聚类、熵权法和TOPSIS方法, 本文测度并比较了各省的数字经济发展水平, 发现各省份的发展水平排名相对稳定, 但也存在显著的省际差异, 特别是在内陆和沿海地区。研究结果揭示了数字经济发展的地域差异性, 并为相关政策的制定和数字经济的持续发展提供了理论支持和实践指导。

关键词

数字经济, 综合评价, 主成分分析, K-Means聚类, 熵权法, TOPSIS计算

Comprehensive Evaluation of National Digital Economy Development

—An Analysis Based on Provincial Panel Data

Zehua Lin, Haidong Liu, Shu Li

College of Science, Tianjin University of Commerce, Tianjin

Received: Sep. 23rd, 2023; accepted: Oct. 14th, 2023; published: Oct. 30th, 2023

Abstract

This study explores the development of the digital economy in China and constructs a comprehensive evaluation system based on provincial panel data from 2013 to 2020. By employing Principal Component Analysis, K-Means clustering, Entropy Weight Method, and TOPSIS method, this paper measures and compares the digital economic development levels of various provinces, finding relatively stable rankings but also significant inter-provincial differences, especially between in-

land and coastal regions. The research results reveal the regional disparities in the development of the digital economy and provide theoretical support and practical guidance for the formulation of related policies and the sustained development of the digital economy.

Keywords

Digital Economy, Comprehensive Evaluation, Principal Component Analysis, K-Means Clustering, Entropy Weigh Method, TOPSIS Calculation

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

(一) 研究背景

1、不断做强做优做大我国数字经济

党的十八大以来，党中央高度重视发展数字经济，将其上升为国家战略。十九届五中全会提出，要发展数字经济，推进数字产业化与产业数字化，推动数字经济和实体经济深度融合，打造具有国际竞争力的数字产业集群。2021年，习近平总书记在致世界互联网大会乌镇峰会的贺信中指出：要激发数字经济活力，增强数字政府效能，优化数字社会环境，构建数字合作格局，筑牢数字安全屏障，让数字文明造福各国人民。

2、各省系列前沿政策举措促发展

为促进数字经济的发展，各省份也陆续发布系列前沿政策。2022年3月22日，湖北省首次发布《湖北省数字经济发展白皮书》，通过采集全省及各市州2018~2020年数字经济相关数据，梳理全省数字经济发展形势、总体情况，有针对性地提出了政策建议。4月26日，江西省推出18条举措，对全省数字经济领域市场主体实行2年发展“包容期”管理，积极营造数字经济发展一流生态，大力推进数字经济做优做强。

(二) 研究目的与意义

1、经济发展在不同程度上受到新冠肺炎疫情的影响，后疫情时代下复苏经济迫在眉睫，而数字经济的发展已是推动经济复苏的核心。

2、基于2013~2020年全国31个省份的数据，构建科学的指标体系，测算各省份数字经济发展情况，从时空维度展开综合评价。

3、本文创新地使用主成分分析、聚类和熵权法等模型，定性定量对各省份的数字经济发展评估，针对建模结果提出可行性建议。

本文研究框架，如图1所示。

2. 文献综述

(一) 基于研究发文量和主题分布的分析

如图2所示，首先通过中国知网检索“数字经济”得到2011~2022年(截止到2022年5月24日)国内相关研究的发文数量，可以看出这一时期内发文数量明显突变时间为2016年。在2011~2015年，每年发文量保持在100篇以内，之后发文量呈现直线增长趋势，2020年发文量达到4787篇，2021年创造年度最高8500篇，而2022年仅前5个月就达到了4023篇之多。由此可见，近年来数字经济已成为我国学术界关注的一大热点，表明了国内学者对数字经济的重视和数字经济发展的的重要性。

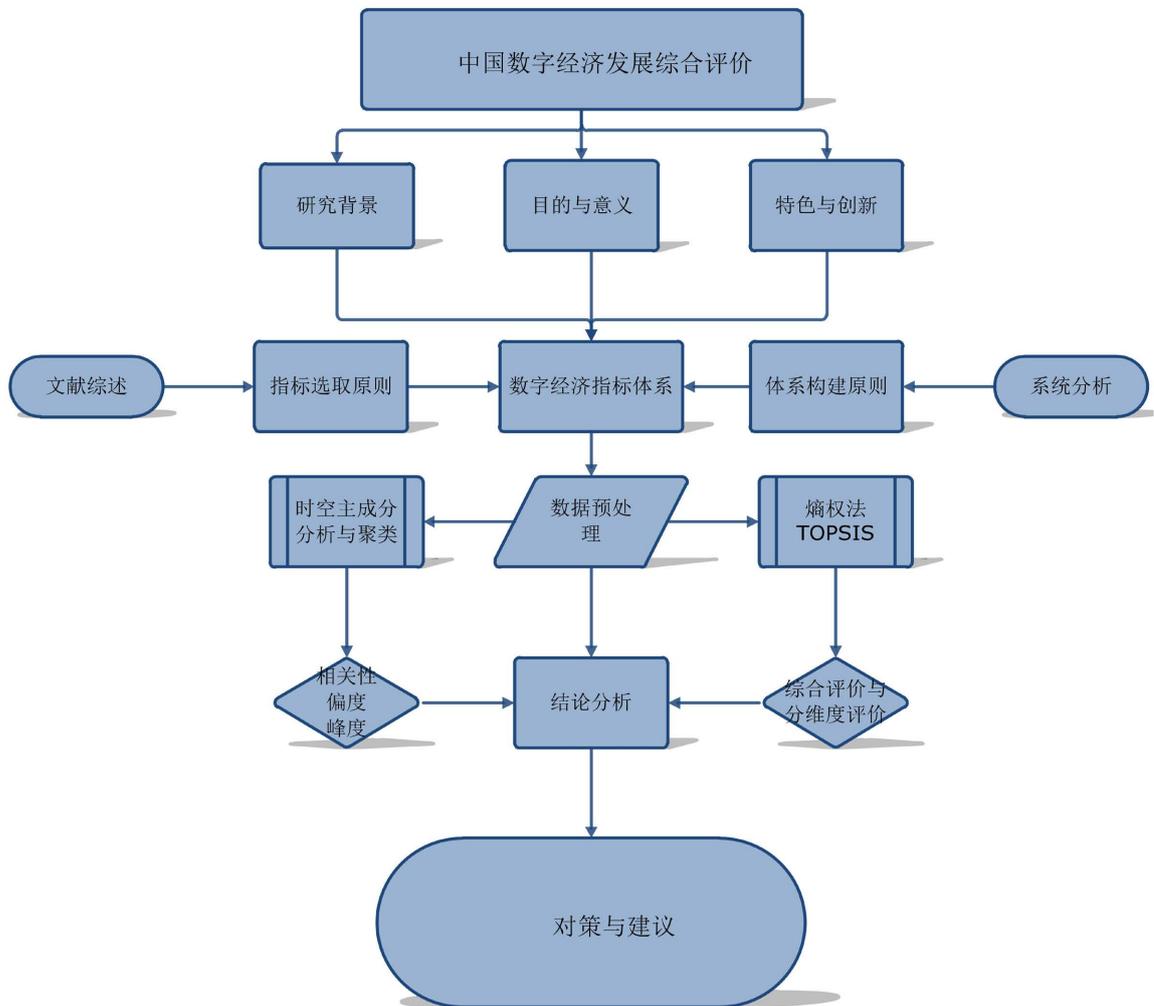


Figure 1. Research framework
图 1. 研究框架

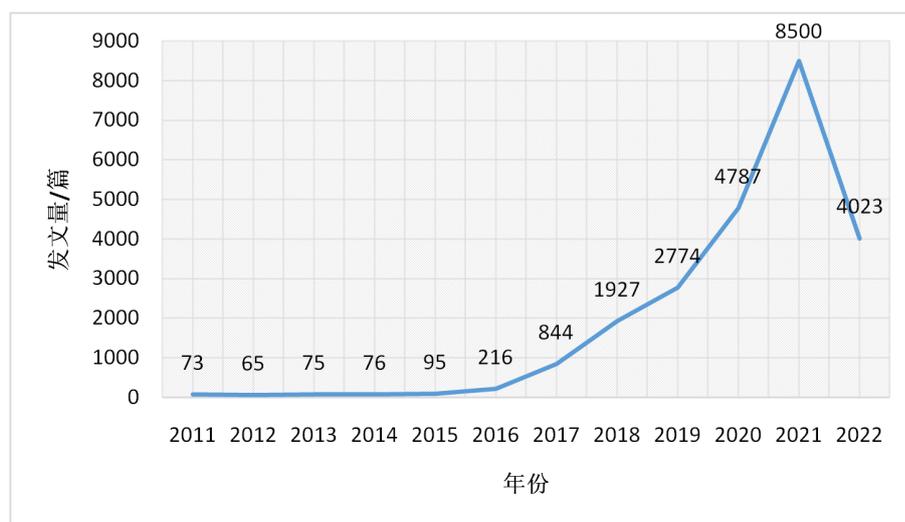


Figure 2. The number of publications on domestic digital economy research from 2011 to 2020
图 2. 2011~2020 年国内数字经济研究发文量

其次，从“数字经济”相关的文章的关键词频数出发，分析研究的主题分布，统计出发文数量排名前 10 的主题如图 3 所示。明显看出“数字经济”主题对应发文量遥遥领先，超过了其余 9 个主题的发文量，达到 5630 篇；而数字化转型、高质量发展等主题依次递减；而由于部分高新技术和新领域，如工业互联网、区块链和数字中国等是近几年才出现在大众的视野中，因此这些主题相关的文献发文量相对偏少，但不可否认的是各类技术和新领域的诞生对数字经济的高质量发展大有裨益。

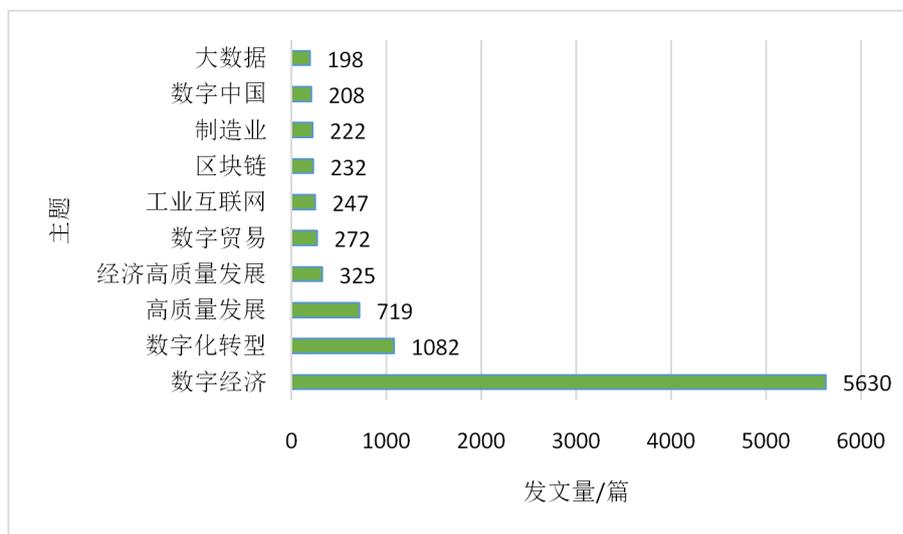


Figure 3. Distribution of research themes on domestic digital economy

图 3. 国内数字经济研究主题分布

(二) 理论方法文献综述

20 世纪 90 年代，Tapscott (1996)首次提出数字经济，他表示数字经济刻画的是一个充分利用 ICT 技术的经济体系，包括电子商务和基础设施等[1]。洪兴建(2019)考虑到学术界和政府部门的出发角度存在差异，因此分别从统计核算、全流程、驱动力和显著特征等出发，界定了数字经济的内涵[2]。当前，根据国内外政府部门和学术界的研究，数字经济的测度大致分为直接法和对比法两类，前者是在界定范围内度量一定区域的数字经济规模，后者是构建综合评价体系评价不同区域的数字经济发展情况，具体如表 1 所示。

Table 1. Classification of measurement methods for digital economy

表 1. 数字经济测度方法分类

测度方法	相关机构	文献/指数
直接法	美国商务部数字经济咨询委员会(DEBA)	-
	中国信息通信研究院	-
对比法	经济合作与发展组织	《数字经济展望 2017》
	欧盟委员会	《数字经济与社会指数 DESI2018》
	中国信息通信研究院	数字经济指数(DEI)
	赛迪顾问	中国数字经济指数(OEDI)

陈梦根等(2020)指出，数字产业化部分的增加值为国民经济核算体系中的不同行业信息产业增加值总和，可借助 KLEMS 模型测算[3]。李英杰等(2020)综合数字经济的外延与内涵，分别从数字产业化和产

业数字化构建数字经济发展量化体系, 根据熵值法测度了中国 2010~2018 年数字经济发展形势, 另外用灰度预测模型预测 2019~2028 年的数字经济进展[4]。许宪春等(2020)在梳理数字经济发展后构建了核算体系, 确定数字经济产品, 筛选出数字经济产业, 对 2007~2017 年中国数字经济的各项指标进行测算[5]。金星晔等(2020)以关键要素的生产应用为基础, 重新定义数字经济, 重点界定产业核算范围, 对数字经济产业活动分类, 讨论了不同产业的核算方法[6]。

韩兆安等(2021)通过 Kernel 密度估计和 Dagum 基尼系数评估法, 测算中国各省数字经济规模[7]。陈亮等(2021)参考国内外对数字经济分类的经验, 制定出符合中国国情的数字经济分类框架, 对 2012 年和 2017 年的规模进行测算[8]。李洁等(2021)为针对我国数字经济规模进行科学测算, 同时注重方法和框架具有国际适用性, 运用投入产出模型针对中国 2012 年、2017 年和 2018 年的数字经济规模测算[9]。陈梦根等(2022)利用投入产出列表数据, 结合数字经济内涵和增长核算方法, 创立数字经济规模和全要素生产率测算体系, 考察中国数字经济的结构特征[10]。

3. 数据与预处理

(一) 数据来源与变量说明

查阅大量文献后, 参考各界对数字经济发展指标体系的构建流程, 本文基于指标的科学性、系统性和可比性, 选取了 4 个维度、共 19 个二级指标构建数字经济发展指标体系[11], 见表 2。由于覆盖 31 个

Table 2. Construction table of indicator system for digital economy development

表 2. 数字经济发展指标体系构建表

一级指标	二级指标(单位)	符号
智能基础设施	光缆长度(公里)	X_1
	移动电话基站数(万个)	X_2
	移动电话普及率(每百人部数)	X_3
	互联网宽带接入端口数(万个)	X_4
	互联网域名数(万个)	X_5
数字金融发展	数字金融覆盖广度	X_7
	数字金融使用深度	X_8
	数字金融数字化程度	X_9
	网上移动支付水平	X_{10}
科研创新能力	规模以上工业企业 R&D 人员折合全时当量(人年)	X_{11}
	规模以上工业企业 R&D 经费支出(万元)	X_{12}
	规模以上工业企业 R&D 项目(课题)数(项)	X_{13}
	专利申请数(件)	X_{14}
	发明专利申请数(件)	X_{15}
	专利申请授权数(件)	X_{16}
信息服务建设	软件业收入(万元)	X_6
	信息服务业从业人数(万人)	X_{17}
	信息服务业产值(亿元)	X_{18}
	电信业务量(亿元)	X_{19}

省份 2013~2020 年的面板数据, 可全面反映各省份数字经济的发展。其中 $X_1 \sim X_7$ 查自国家统计局和《中国统计年鉴》, $X_{11} \sim X_{16}$ 、 $X_{17} \sim X_{19}$ 来自《中国第三产业统计年鉴》和《中国信息产业年鉴》、 $X_7 \sim X_{10}$ 来源于北京大学数字普惠金融指数。

(二) 数据预处理

1、缺失值处理

本文“信息服务业产值”和“信息服务业产值”指标存在少量缺失值, 基于回归插值法, 根据同一省份的现有数据, 对缺失值进行估算。具体思路为构建自变量 $X_i (i=1,2,\dots,m)$ 与目标变量 Y 的关系, 则第 k 个缺失值的插补估算值可由公式(1)计算[12], 其中 ζ 指随机因素。

$$y_k = \alpha_0 + \sum_{i=1}^m \alpha_i X_{ik} + \zeta_k \tag{公式(1)}$$

2、异常值处理

异常数据可能来自数据本身、存储或转换过程, 可删除记录或视为缺失值。设各指标在各年取值为 $X_{ij} (i=1,2,\dots,31; j=1,2,\dots,19)$, 算出各指标平均值 \bar{X}_i 及取值的误差 $V_{ij} = X_{ij} - \bar{X}_i$ 和指标标准差 σ_i 。

若某测量值 X_b 的剩余误差 $V_b (1 \leq b \leq n)$ 满足 $|V_b| = |X_b - \bar{X}_i| > 3\sigma$, 则认为 X_b 应予修正。本文指标“数字金融覆盖广度”和“网上移动支付水平”存在异常值, 采取该准则对数据进行处理。

3、数据标准化

不同指标有不同的单位, 会影响分析结果。本文为消除不同量纲的影响, 采取归一化方式对数据标准化, 使得原始数据结果映射到 0~1 之间, 即公式(2):

$$x_{ij}(t) = \left[\frac{x_{ij}(t)^* - \min x_{ij}(t)^*}{\max x_{ij}(t)^* - \min x_{ij}(t)^*} \right] \tag{公式(2)}$$

其中,

$$\left\{ x_{ij}(t)^* \right\}_{n \times p \times d} \quad (i=1,2,\dots,31; j=1,2,\dots,19; t=1,2,\dots,8)$$

为多指标面板数据, $x_{ij}^*(t)$ 表示第 i 个样本在 t 时刻的第 j 个指标的原始观测值。 $\left\{ x_{ij}(t)^* \right\}_{n \times p \times d}$ 为标准化结果, $\max x_{ij}(t)^*$ 和 $\min x_{ij}(t)^*$ 分别为指标的最大值和最小值。

4. 基于时空主成分分析和聚类的评价

(一) 主成分分析

1、相关性分析

设两变量为 x_i 和 x_j , 按照公式(3)计算相关系数 r , 表示变量 x_i 与 x_j 间的相关性。

$$r(x_i, x_j) = \frac{\text{Cov}(x_i, x_j)}{\sqrt{\text{Var}(x_i)\text{Var}(x_j)}} \tag{公式(3)}$$

19 个指标在不同时期的相关系数见表 3 (详细见附录), 表 3 仅展示变量 x_1 与其他 18 个变量之间的相关性。同一变量在不同年份的相关性均较强, 可通过显著性检验, 对各指标间的相关系数展开分析[13]。

以指标光缆长度为例分析为例, 其与移动电话基站数之间、互联网宽带接入端口数、规模以上工业企业 R&D 项目(课题)数、电信业务量之间的相关系数均为正。光缆越长, 相应的依靠光缆发展的业务越强劲, 光缆长说明发展电信业务具备的基本条件充分。其他指标也可作类似分析。

Table 3. Secondary indicator correlation coefficient table
表 3. 二级指标相关系数表

变量	2013	2014	2015	2016	2017	2018	2019	2020
1,2	0.91	0.91	0.92	0.93	0.89	0.87	0.88	0.87
1,3	0.04	0.02	0.08	0.01	-0.09	-0.08	-0.03	-0.19
1,4	0.88	0.89	0.93	0.92	0.92	0.9	0.89	0.88
1,5	0.35	0.38	0.37	0.35	0.2	0.3	0.38	0.49
1,6	0.54	0.56	0.58	0.57	0.48	0.4	0.36	0.32
1,7	0.22	0.23	0.25	0.22	0.17	0.13	0.14	0.15
1,8	0.41	0.23	0.39	0.37	0.33	0.32	0.32	0.29
1,9	-0.13	-0.27	-0.52	-0.11	0.27	0.37	0.37	0.36
1,10	0.35	0.4	0.41	0.46	0.46	0.42	0.38	0.34
1,11	0.81	0.82	0.84	0.84	0.79	0.73	0.72	0.72
1,12	0.76	0.76	0.78	0.8	0.76	0.74	0.73	0.73
1,13	0.79	0.83	0.84	0.83	0.78	0.76	0.75	0.74
1,14	0.8	0.82	0.82	0.8	0.74	0.69	0.66	0.66
1,15	0.68	0.69	0.71	0.73	0.7	0.66	0.57	0.56
1,16	0.8	0.81	0.82	0.78	0.69	0.66	0.64	0.66
1,17	0.41	0.46	0.48	0.51	0.48	0.42	0.4	0.42
1,18	0.41	0.4	0.44	0.45	0.41	0.3	0.29	0.3
1,19	0.76	0.78	0.81	0.83	0.78	0.81	0.83	0.83

2、基于时空的主成分分析

本文共使用基于时空的主成分分析，即两次主成分分析评价中国数字经济的发展状况。第一次是在不同年份分别进行，结合特征值大于 1 的原则提取主成分，通过回归获得主成分综合得分[14]。第二次是

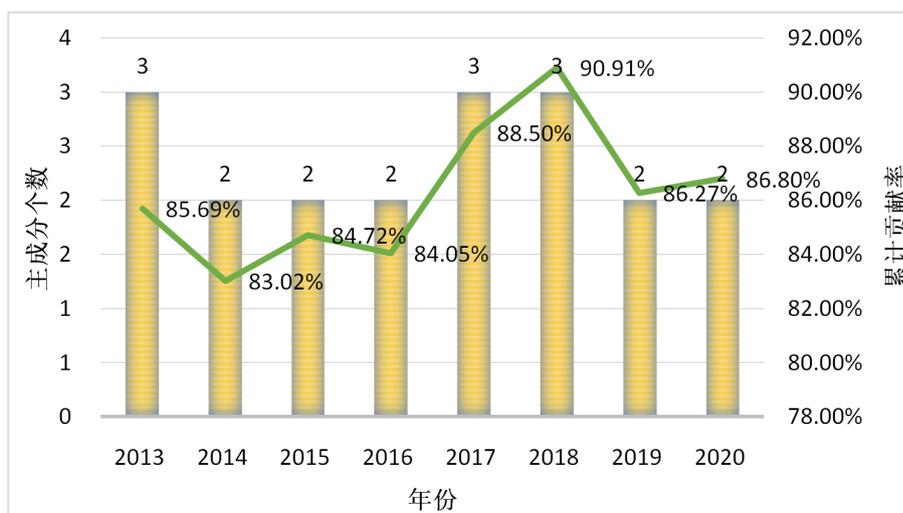


Figure 4. Principal component of the first principal component analysis
图 4. 第一次主成分分析主成分

基于前面各次分析的主成分综合得分，筛选第二次分析的主成分，通过回归获得综合得分，最终运用该结果评价 31 个省份的数字经济发展。

在固定年份，将 31 个省份的 19 个指标作为样本观测值，计算相关系数结果显示整体呈现较强相关，适合主成分分析[15]。根据特征值大于 1 的原则，选取各次分析的主成分个数和累积贡献率见图 4。运用回归方法得到各省份的综合得分进行排名的结果见表 4。

Table 4. Results of the first principal component analysis

表 4. 第一次主成分分析结果

地区	2013	2014	2015	2016	2017	2018	2019	2020
广东	1	1	1	1	1	1	1	1
北京	3	3	3	3	4	4	2	2
江苏	2	2	2	2	2	3	4	3
浙江	4	4	4	4	3	2	3	4
上海	5	5	5	5	5	5	5	5
山东	6	6	6	6	7	7	7	6
福建	7	7	7	7	6	6	6	7
湖北	10	11	9	9	9	8	8	8
河南	12	12	13	12	11	11	9	9
四川	9	8	8	8	8	10	10	10
安徽	13	14	11	10	10	9	11	11
天津	11	10	10	11	12	12	12	12
湖南	17	16	15	15	14	15	13	13
陕西	15	15	16	16	16	16	14	14
重庆	14	13	14	14	13	13	15	15
江西	22	27	18	18	17	14	17	16
河北	16	17	17	17	18	18	16	17
辽宁	8	9	12	13	15	17	18	18
广西	26	20	20	20	19	19	19	19
山西	21	21	22	23	21	22	21	20
海南	20	18	21	19	23	20	20	21
云南	24	23	23	24	24	21	22	22
贵州	27	28	27	28	25	23	23	23
吉林	25	22	24	21	26	24	25	24
黑龙江	19	19	19	22	22	25	24	25
内蒙古	18	26	26	27	20	28	26	26
新疆	23	25	25	25	27	27	27	27
甘肃	28	30	30	29	30	30	28	28
宁夏	29	24	28	26	29	29	29	29
西藏	31	31	29	30	28	26	30	30
青海	30	29	31	31	31	31	31	31

广东在数字经济发展方面稳居第一；北京、江苏、浙江和上海排名较稳定，维持在 2~5 名；河南、湖南、江西、广西、贵州等省份从 2013 年到 2020 年进步较大；而黑龙江、辽宁、内蒙古、新疆等省份退步较大；甘肃、宁夏、西藏、青海排名时常垫底，其他省份的变化不太明显，趋于平稳。

Table 5. Skewness and kurtosis of the comprehensive score of the first principal component

表 5. 第一次主成分综合得分偏度与峰度

年份	2013	2014	2015	2016	2017	2018	2019	2020
偏度	1.50	1.45	1.41	1.47	1.20	1.21	1.42	1.38
峰度	1.28	1.00	0.95	1.16	0.56	0.58	1.23	1.12

第一次主成分综合得分的偏度与峰度结果见表 5，峰度最大为 1.28，说明全国数字经济的发展不受极端值的影响。偏度均大于零，说明数据分布右偏，中位数小于平均数，即全国有一半以上省份数字经济低于全国平均水平。第二次主成分分析运用第一次主成分分析的综合得分，分析前先进进行相关分析，结果如图 5 所示。可知相关关系很强，说明 8 年间 31 个省份的数字经济发展较稳定，这得益于国内稳定的政治经济环境，可再一次进行主成分分析。

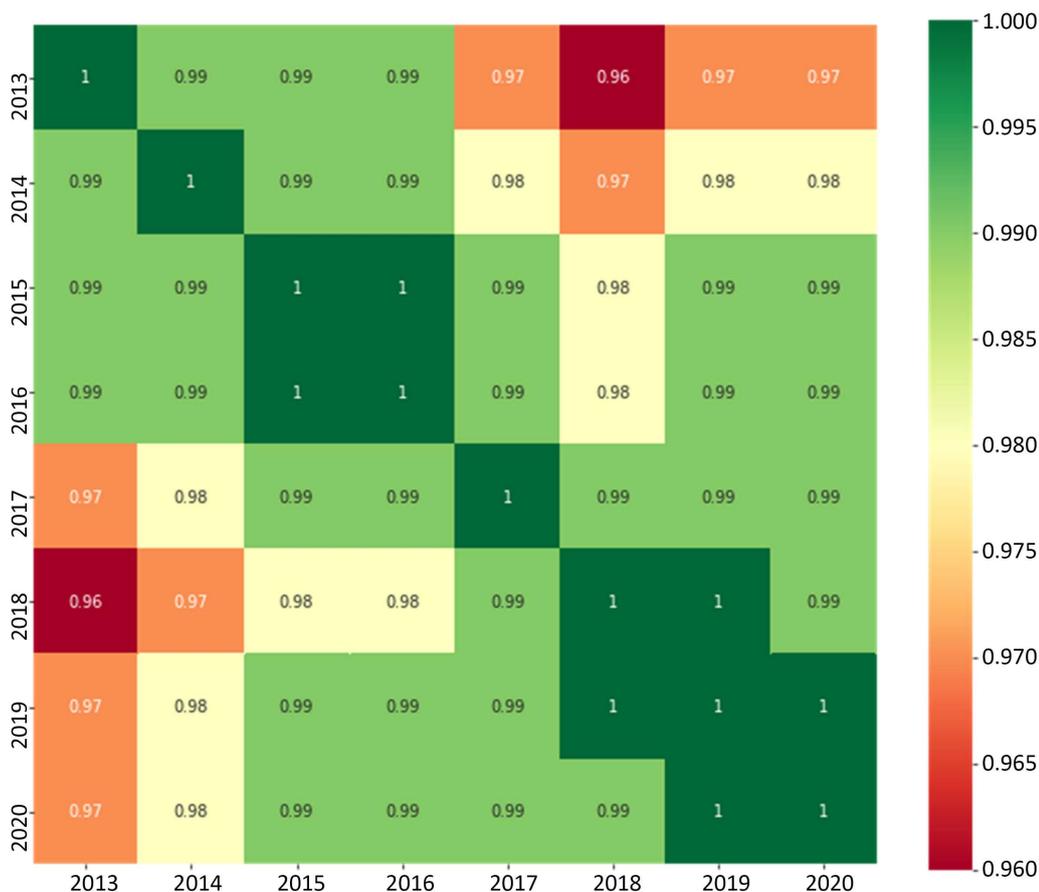


Figure 5. Heatmap of correlation coefficients for the comprehensive score of the first principal component

图 5. 第一次主成分综合得分相关系数热力图

根据特征值大于 1 的原则，在 2013~2016 年、2017~2020 年和 2013~2020 年各选取一个主成分，进行第二次主成分分析。考虑了时间间隔的影响，重新评价各省份的数字经济发展，排名结果见图 6。结合 2013~2020

年的表现, 可知在三个时段, 各省份排名浮动不大, 广东、北京、江苏和浙江始终排名前四; 甘肃、西藏和青海排名一直位于后四位。2017~2020 年较 2013~2016 年排名进步超过 3 位的有: 湖南、江西、贵州和西藏, 2017~2020 年较 2013~2016 年排名下降超过 3 位的有: 辽宁、黑龙江与宁夏。其他省份的变化比较平缓。2013~2020 年整体来看, 变化较大的省份为上述提到的七个省份, 其他省份的变化也较为平缓。

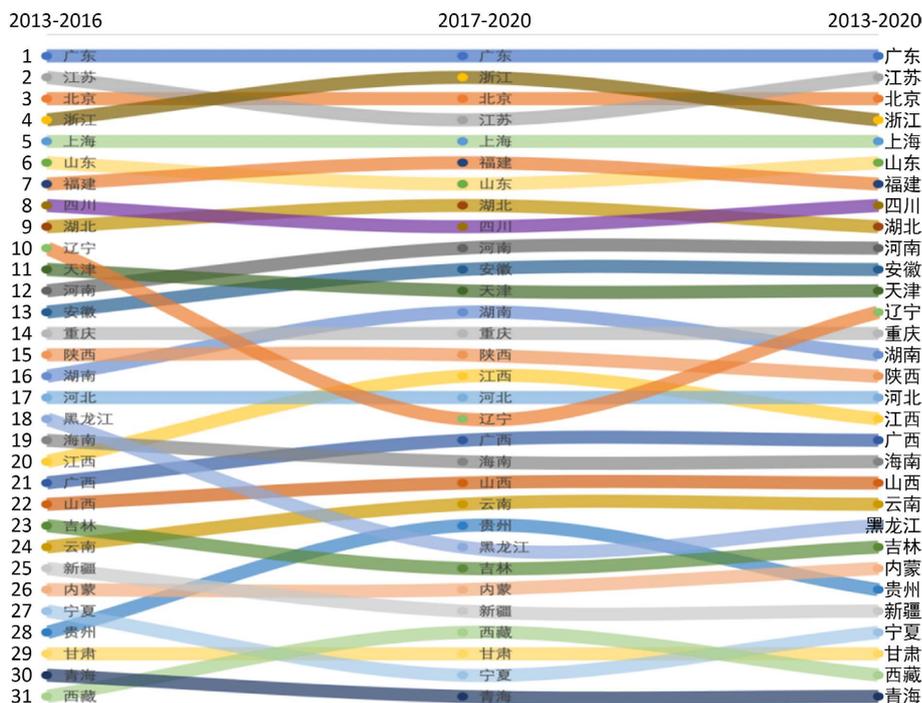


Figure 6. Results of the second principal component analysis
图 6. 第二次主成分分析结果

对各时段综合得分进行偏度与峰度分析, 结果见表 6。按照时间来看, 偏度值在下降, 说明全国各地区的数字经济发展水平在不断上升且逐渐平缓, 即全国各地区小于全国数字经济平均发展水平的地区数量逐渐减少, 这是区域均衡发展的趋势。峰度也能说明类似结果。考察各省在不同年份排名的 Spearman 等级相关系数, 结果见图 7, 发现相关性都较强, 其中 2018 年与 2013 年、2014 年的等级相关系数最小为 0.91, 表明我国经济系统比较稳定。

Table 6. Skewness and kurtosis of the comprehensive score of the second principal component
表 6. 第二次主成分综合得分偏度与峰度

年份	2013~2016	2017~2020	2013~2020
偏度	1.46	1.30	1.39
峰度	1.09	0.84	0.99

根据 31 个省份在不同阶段数字经济的综合得分, 利用 python 绘制综合得分热力图以展示数字经济发展的整体情况。(以 2013~2016 年和 2017~2020 年地图为例, 见图 8、图 9)由图 8 可知, 数字经济发展存在区域差距, 沿海与内陆地区的差异明显。广东省的数字经济发展最好, 综合得分为 2 以上。数字经济发展较好地区一般集中于东南沿海地区, 当地经济发展水平也相对更高; 其他地区的发展水平较为接近, 都属于数字经济发展较不发达地区。这反映了我国数字经济发展存在不平衡问题, 在基础设施建设等方面的差距不利于数字经济的发展, 造成了内陆地区发展受阻。

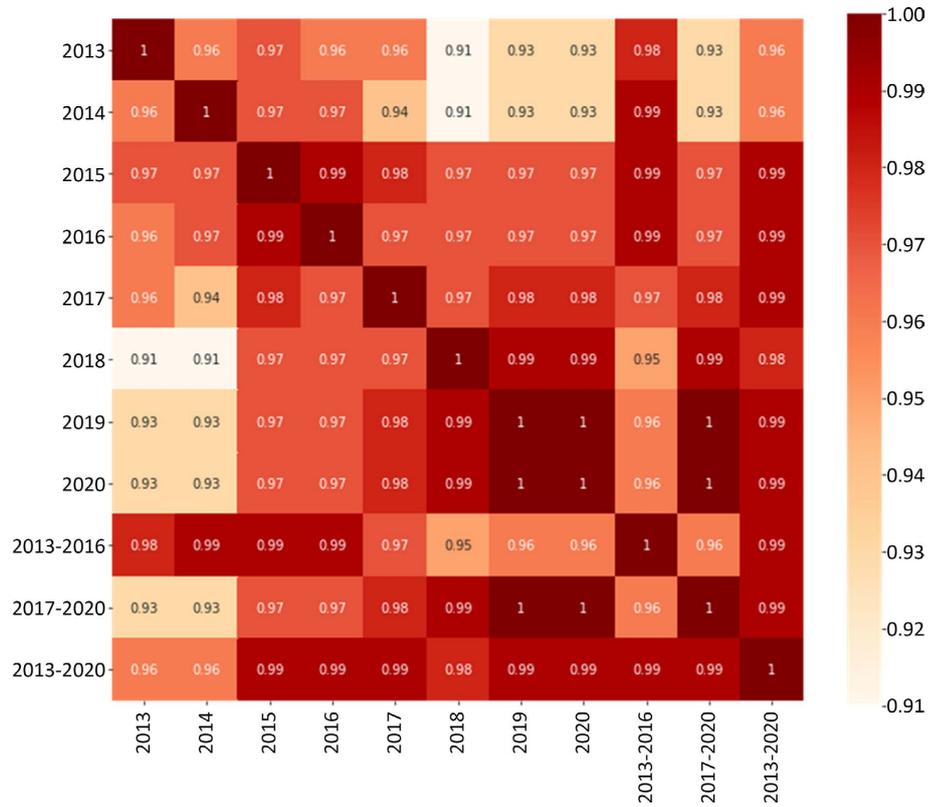


Figure 7. Heatmap of spearman rank correlation coefficients
图 7. Spearman 等级相关系数热力图

2013-2016年各省份综合得分

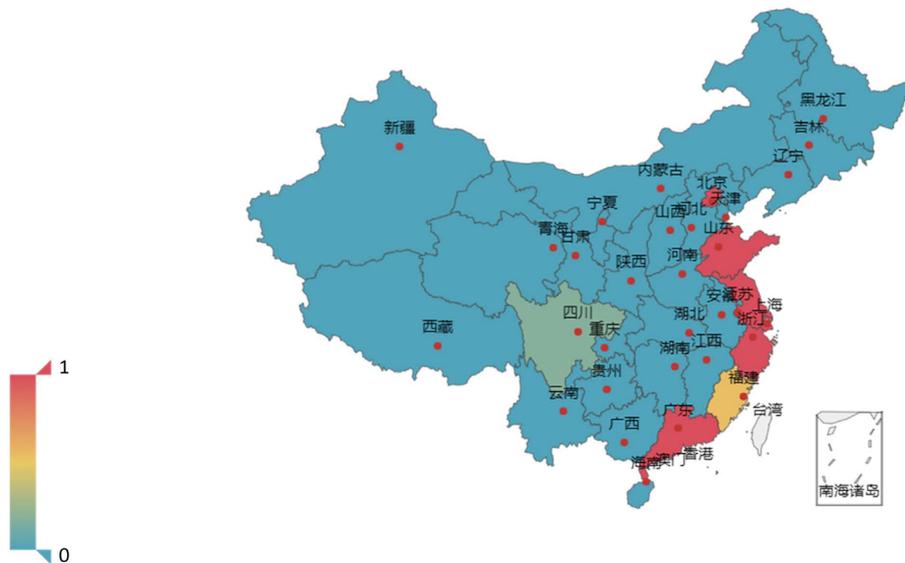


Figure 8. Heatmap of comprehensive scores for each province from 2013 to 2016
图 8. 2013~2016 年各省份综合得分热力图

由图 9 可知，与 2013~2016 年相比，湖北、四川等省份数字经济进步明显，综合得分在较发达地区和发达地区整体呈逐年递增趋势。区域数字经济发展差距主要表现在内陆与沿海地区的差距。从东部往

西部差距逐渐递减，且数字经济发达程度的区位与我国宏观经济发展存在高度一致。要想缩小区域数字经济的差距，推动欠发达地区数字经济的发展是当务之急[16]。

2017-2020年各省份综合得分

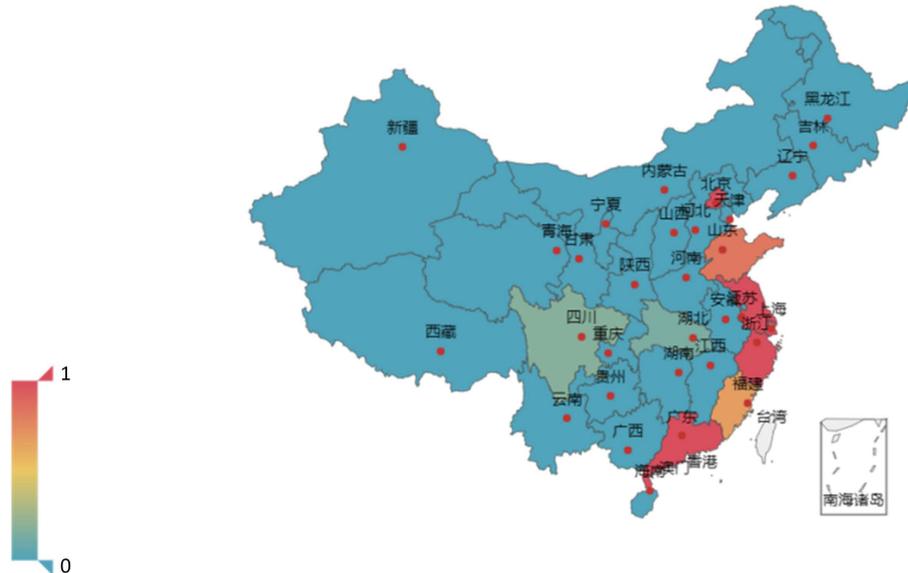


Figure 9. Heatmap of comprehensive scores for each province from 2017 to 2020
图 9. 2017~2020 年各省份综合得分热力图

(二) K-Means 聚类

1、算法说明

K 代表类别数量，Means 代表每个类别内的均值，所以 K-Means 算法又称 K 均值算法[17]。它能根据数据样本间的相似性，将数据样本自动分为 K 个簇，相似样本会尽可能被聚到一个簇。如果几个样本相似，就会被聚为一类，形成一个个簇。每个簇有一个中心位置，被称作为质心。

算法目的是选择合适质心，使得在每个簇内各样本与质心之间距离尽量小，保证簇内样本相似性较

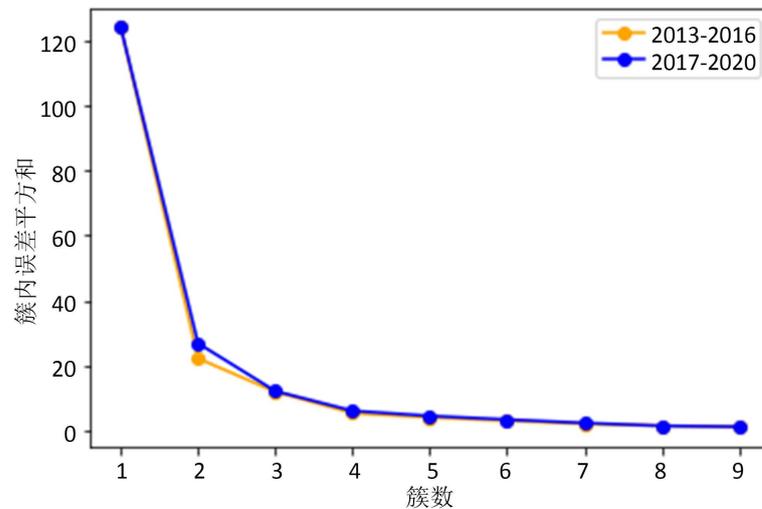


Figure 10. Within-cluster sum of squares error for k-means clustering
图 10. K-means 聚类簇内误差平方和

高，不同簇中数据尽量不同。要确定应该把样本分为几类合适，可以借助肘部法则，即取不同的 K 值，多次进行聚类，并获取每一次的簇内误差平方和(公式(4))，然后将结果进行绘制；随着 K 值增大，质心变多，SSE 会随之降低，当下降幅度趋向于缓慢时，认为是最佳 K 值[18]。

$$SSE = \sum_{i=1}^k \sum_{j=1}^{m_i} \|x_j - \mu_i\|^2 \tag{公式(4)}$$

其中， k 为簇的数量， m_i 为第 i 个簇含有的样本数量， μ_i 为第 i 个簇的质心， $\|x_j - \mu_i\|$ 为第 i 个簇中每个样本 x_j 与质心 μ_i 的距离。

2、实证分析

根据 2013~2016 年和 2017~2020 年的综合得分，运用 K-Means 聚类对 31 个省份进行聚类，取不同的 K 值聚类得到每次簇内误差平方和，绘制结果见图 10。根据肘部法则可知在 $K = 4$ 时，增加 K 值能让聚类的效果显著提升，故可将省份数字经济分为高水平、较高水平、中等水平和低水平四类，见表 7。

Table 7. Cluster results of digital economy in various provinces
表 7. 各省份数字经济聚类结果

类别	2013~2016 年	2017~2020 年
高水平	广东, 北京, 江苏, 浙江	广东, 北京, 江苏, 浙江
较高水平	上海, 山东, 福建	上海, 山东, 福建
中等水平	湖北, 河南, 四川, 安徽, 天津, 湖南, 陕西, 重庆, 河北, 辽宁	湖北, 河南, 四川, 安徽, 天津, 湖南, 陕西, 重庆, 江西, 河北, 辽宁
低水平	江西, 广西, 山西, 海南, 云南, 贵州, 吉林, 黑龙江, 内蒙古, 新疆, 甘肃, 宁夏, 西藏, 青海	广西, 山西, 海南, 云南, 贵州, 吉林, 黑龙江, 内蒙古, 新疆, 甘肃, 宁夏, 西藏, 青海

5. 基于熵权法和 TOPSIS 计算的测度与分析

(一) 理论基础

1、熵权法计算[19]

1) 对原始矩阵 $X = (x_{ij})_{m \times n}$ 进行线性 Max-Min 标准化:

$$Y_{ij} = \begin{cases} \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})}, x_{ij} \text{ is } + \\ \frac{\max(x_{ij}) - x_{ij}}{\max(x_{ij}) - \min(x_{ij})}, x_{ij} \text{ is } - \end{cases} \tag{公式(5)}$$

由于所有的指标均为正向指标，所以对数据进行标准化时采用相同的方式。

2) 计算第 j 项指标在第 i 年的数值占该指标比重 P_{ij} :

$$P_{ij} = \frac{Y_{ij}}{\sum_{i=1}^m Y_{ij}} \tag{公式(6)}$$

3) 计算第 j 项指标熵值 E_j 。当 $P_{ij} = 0$ 时， $P_{ij} \ln P_{ij} = 0$,

$$E_j = -\frac{1}{\ln m} \sum_{i=1}^m P_{ij} \ln P_{ij}, (j = 1, 2, \dots, n) \tag{公式(7)}$$

4) 计算第 j 项指标差异系数 G_j :

$$G_j = 1 - E_j \tag{公式(8)}$$

5) 计算第 j 项指标的权重 W_j :

$$W_j = \frac{G_j}{\sum_{j=1}^n G_j}, (j=1, 2, \dots, n) \tag{公式(9)}$$

2、TOPSIS 计算

6) 由标准化矩阵和各指标权重可得加权标准化矩阵:

$$Z_{ij} = W_j \times Y_{ij} \tag{公式(10)}$$

7) 计算正负理想解:

$$I_j^+ = \max(z_{1j}, z_{2j}, \dots, z_{mj}) \tag{公式(11)}$$

$$I_j^- = \min(z_{1j}, z_{2j}, \dots, z_{mj}) \tag{公式(12)}$$

8) 计算各指标与正负理想解间欧氏距离:

$$D_i^+ = \sqrt{\sum_{j=1}^n (i_j^+ - z_{ij})^2} \tag{公式(13)}$$

$$D_i^- = \sqrt{\sum_{j=1}^n (i_j^- - z_{ij})^2} \tag{公式(14)}$$

9) 计算综合评价价值, 显然 $0 \leq C_i \leq 1$, 且 C_i 越大 D_i^+ 越小, 即越接近最大值:

$$C_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{公式(15)}$$

(二) 数字经济发展评价结果

1、综合评价

数字经济综合评价指标 C_i 在一定程度上反映了各省数字经济的发展水平, 其值越接近 1, 表明该省份的经济发展水平越高, 数字经济活力越强[20]。本文按照 C_i 对全国 31 省份进行排序, 结果如图 11 所示。

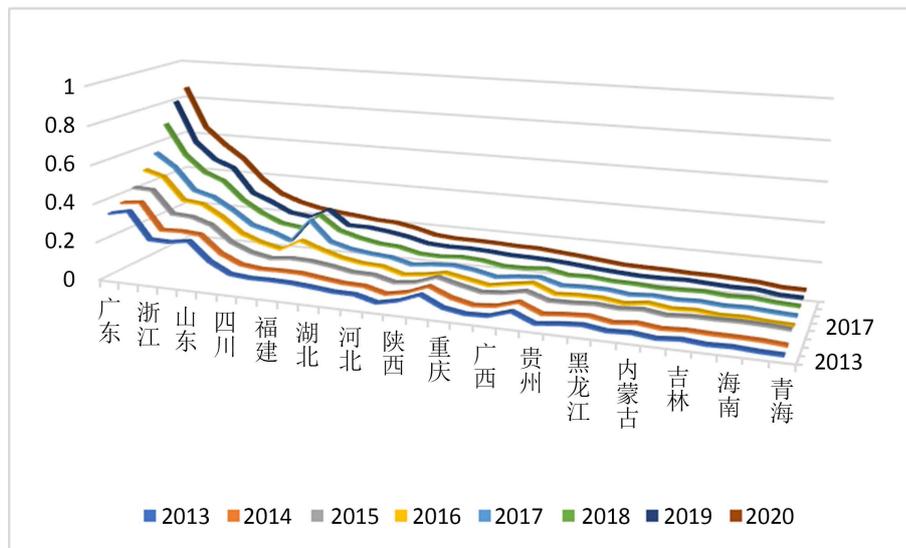


Figure 11. Ranking of comprehensive evaluation of digital economy in various provinces
图 11. 各省份数字经济综合评价排序

由图 11 可知, 全国数字经济整体水平稳步提高, 但各省份经济发展存在严重的地域不平衡性。发展形势较好的地区主要是东南沿海省份如广东、江苏及浙江等; 中部省份江西、山西、陕西等省份的数字经济发展指数较为平均, 处于缓慢上升的趋势中, 没有拉开较大的发展差距, 其中福建省在 2015~2017 发展速度较快, 但近几年增速下滑; 西北省份如青海、西藏、宁夏等省的数字经济发展相较于其他省份明显处于劣势, 发展速度较缓慢, 其中西藏数字经济发展水平在 2016 年之后发展较快, 反超甘肃、宁夏。

在沿海省份中, 广东的数字经济发展水平遥遥领先, 在 2017 后快速发展, 2018 年达到了 24% 的增长率, 增速大大超过了江苏的 8.9% 和浙江的 14.4%, 发展水平增速突出, 逐渐与排在之后的江苏和浙江拉开差距。相比而言, 同为沿海省份的山东、福建, 天津相比江浙粤发展动力欠缺, 发展速度较慢, 在 2020 年仅分别达到了 0.38507, 0.237196, 0.136356, 相比一些内陆省份犹有不及。

本文根据数字经济评价指标将各省发展程度分为了以下三个层次, 第一梯队为广东、江苏、浙江、北京、山东五省份, 显著高于平均水平; 第二梯队为发展水平处于中等水平的四川、河南、福建、安徽、湖北五省为代表(发展水平处在均值附近), 第三梯队以明显较为落后的吉林、甘肃、海南、宁夏、青海五省为代表。

2、分维度评价

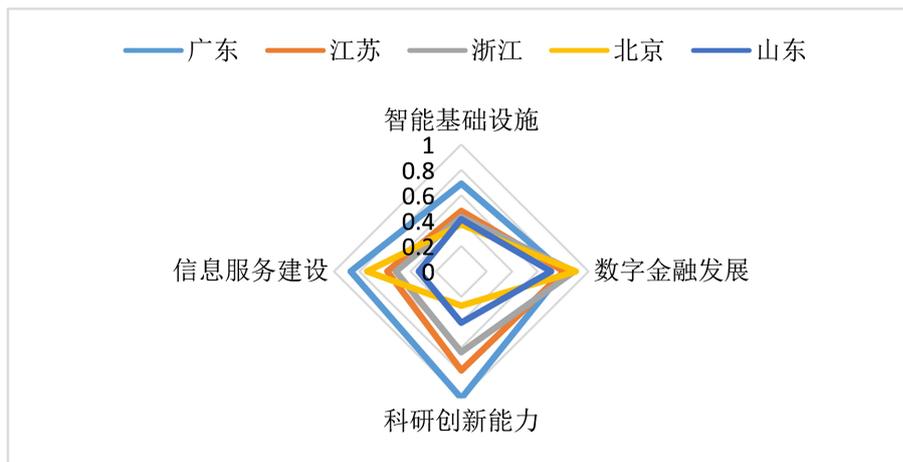


Figure 12. Distribution diagram of secondary indicators for high-level echelon
图 12. 高水平梯队二级指标分布图

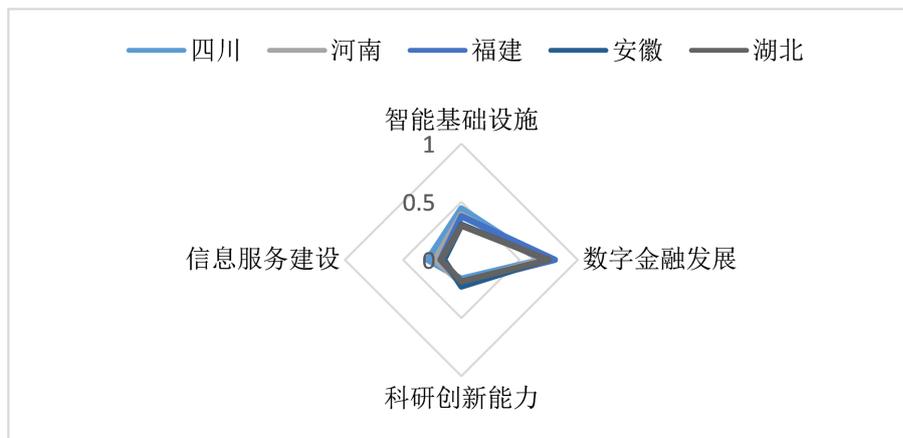


Figure 13. Distribution diagram of secondary indicators for medium-level echelon
图 13. 中等水平梯队二级指标分布图

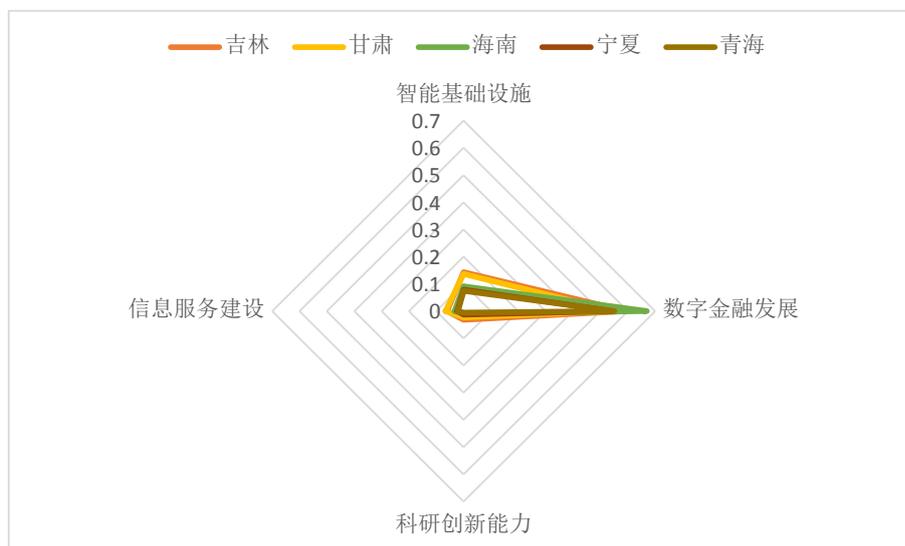


Figure 14. Distribution diagram of secondary indicators for low-level echelon
图 14. 低水平梯队二级指标分布图

在高中低三个层次中，本文以 2020 年为例，分别以 4 个维度：智能基础设施(A)，数字金融发展(B)，科研创新能力(C)，信息服务建设(D)，评价不同层次的地区数字经济发展水平。

根据三个梯队二级指标分布绘制雷达图，如图 12、图 13 和图 14 所示。对比发现：数字经济发展水平较高省份在四个维度均有良好表现，没有明显短板，具有全方位驱动的特征。中等发展水平省份在智能基础设施方面具有较好表现，而在科研创新能力和信息服务建设方面较弱，尤其是前者上存在巨大的短板。低发展水平的省份在四个维度处于落后，但在数字金融方面有较好表现，这与全国范围的数字金融使用深度加深，移动支付水平提高等有关。

6. 政策建议

(一) 政策建议

1、全力推进“东数西算”工程，协同联动东西部

全国大数据中心体系总体布局一体化设计的完成，标志着“东数西算”工程的全面启动。首先，政府部门要同推动“西电东送”、“西气东输”和“南水北调”工程一样，充分发挥我国体制优势，优化资源配置，促进东西部的协同联动，提升资源使用效率。其次，相关部门要根据算力需求和业务对网络的要求，将数据中心自东向西梯度布局、统筹发展，比如人工智能推理、远程医疗等可布局于粤港澳大湾区、京津冀和长三角等算力枢纽，而离线分析、存储备份可优先向西部转移。同时要大力打造一批“东数西算”示范线路，加快完善数据中心产业生态体系，着力于提升 8 个枢纽的影响力和集聚力。

2、加快形成数字经济省域协调发展空间布局

相比内陆省份如青海、贵州和宁夏等省份，广东、浙江和江苏等沿海地区省份在智能基础设施、数字金融发展、科研创新能力和信息服务建设四个维度上全方位领先，显著优势。从习近平总书记坚持全国一盘棋的角度出发，发展水平较高的沿海地区应该对中西部内陆地区开展如“一对一”、“点对点”的帮扶，缓解地区间发展的不平衡。党的十九大提出，我国数字经济应从短平快的发展道路转换为高质量发展，中西部地区应借助这一东风，在十四五期间实现阶段性跳跃。

3、全面提高智能基础设施建设水平

各地区应积极打造互联网智能基础设施，加快建立大数据中心，推动数据的智能化和绿色化发展，

构建高效数据处理系统,提速 5G 网络的部署工作,推进各类感知设施布局,真正形成共建共享的城市物联网。全面促进新技术在传统设施上的使用,加快交通、能源等基建的数字化改造,推进能源互联网建设,响应国家双碳政策,实现公共服务一体化、智慧化。最后,科学规划创新基础设施体系,推动人工智能等高新技术的有机融合,以智能制造、信息安全为核心领域,更快推进产业创新平台的建立,建设智能车间和数字化工厂,提高相关领域的交通设施车联网的改造速度,促进核心系统能力的培养。

4、大力发展数字金融和普惠金融,弥合数字鸿沟

当前仍存在金融服务的数字鸿沟,对包容性发展形成了障碍。中国的互联网普及率在 2020 年年底已达到 73%,但仍存在较大地区差异。第一,这抑制了金融服务的正常获取,尤其是低收入群众等特殊群体理解能力差,使用数字经济的能力偏低;第二,一些小微企业的金融服务获得感受挫,因为它们的数字转型能力较弱,数据开发投入较少,可使用的数字化工具有限,无法储存足量数字化资源;第三,让相关机构获能够滥用金融交易支撑数据,利用算法欺骗消费者,使得金融数据存在安全隐患。相关部门应深入研究,积极促进平台经济健康发展,推动腾讯、京东等大型互联网金融平台不断为人民的生活提供高质量服务。

5、不断提高科研创新能力,以科技兴国

实践证明一个国家必须把握核心技术,才能从根本上保障国家安全。“卡脖子”技术对我们来说,是挑战也是机遇。国家应继续大力推崇自主创新,强化战略科技力量,提高创新效能;明确企业、高校和科研院所的创新主体定位,激发创新活力。

参考文献

- [1] 姚正海,莫晓雪,张琳若,等.基于 CNKI 期刊文献的我国数字经济研究现状与趋势分析[J].经营与管理,2022(5): 154-161.
- [2] 洪兴建.数字经济:内涵、核算与评价[J].中国统计,2019(8): 49-52.
- [3] 陈梦根,张鑫.数字经济的统计挑战与核算思路探讨[J].改革,2020(9): 52-67.
- [4] 李英杰,韩平.中国数字经济发展综合评价与预测[J].统计与决策,2022,38(2): 90-94.
- [5] 许宪春,张美慧.中国数字经济规模测算研究——基于国际比较的视角[J].中国工业经济,2020(5): 23-41.
- [6] 金星晔,伏霖,李涛.数字经济规模核算的框架、方法与特点[J].经济社会体制比较,2020(4): 69-78.
- [7] 韩兆安,赵景峰,吴海珍.中国省际数字经济规模测算、非均衡性与地区差异研究[J].数量经济技术经济研究,2021,38(8): 164-180.
- [8] 陈亮,孔晴.中国数字经济规模的统计测度[J].统计与决策,2021,37(17): 5-9.
- [9] 李洁,张天顶.投入产出分析与中国数字经济规模的测量[J].当代经济管理,2021,43(10): 66-73.
- [10] 陈梦根,张鑫.中国数字经济规模测度与生产率分析[J].数量经济技术经济研究,2022,39(1): 3-27.
- [11] 翟文静,许学军.浙江省数字经济发展水平评价研究[J].科技和产业,2022,22(2): 283-286.
- [12] 宋亮,万建洲.缺失数据插补方法的比较研究[J].统计与决策,2020,36(18): 10-14.
- [13] 贾俊平,何晓群,金勇进.统计学[M].第八版.北京:中国人民大学出版社,2021.
- [14] 陈黎,陈平星.主成分分析与熵值法综合评价云南省乡村振兴发展水平[J].中国商论,2023(18): 161-164.
- [15] 魏艳华,马立平,王丙参.中国八大综合经济区经济发展差异测度与评价[J].数量经济技术经济研究,2020,37(6): 89-108.
- [16] 陈晓东,刘冰冰.数字经济推动区域协调发展的实证研究——基于 2013-2019 年中国省级面板数据[J].经济研究参考,2022(4): 78-97.
- [17] 何峻,黄思刚.基于机器学习语言的我国地区快递发展问题研究[J].物流工程与管理,2021,43(5): 33-35.
- [18] 姜启源,谢金星,叶俊.数学模型[M].第五版.北京:高等教育出版社,2018.
- [19] 贝泓涵,朱良富,孙钰童.绿色港口评价指标体系构建[J].港口科技,2021(11): 20-23.
- [20] 赵新伟,王琦.我国数字经济发展空间演进及地区差距分解[J].统计与决策,2021,37(17): 25-28.