

基于全局主成分-聚类分析的城市创新生态系统评价

——以中国 19 个重点城市为例

蔡 红

(深圳市维度数据科技股份有限公司, 广东 深圳 518000)

摘要:基于 19 个重点城市创新要素实际,构建创新环境、创新支付、创新动能、发展趋势、开放水平 5 个维度 20 个指标的城市创新生态系统评价指标体系,根据 2017—2019 年数据,应用全局主成分模型提取主成分和计算得分,并进行聚类分析。研究发现:创新环境与创新支付影响显著;各城市重视程度提高,但得分差距拉大;城市间差异明显,各有优势。由此建议提升创新环境品质、建立全生态创新系统、培育城市创新动能、促进跨区域创新系统协同发展。

关键词:城市创新生态系统;全局主成分分析;聚类分析;评价;协同发展

中图分类号:F124.3;F124.6 **文献标志码:**A **文章编号:**1671-1807(2022)05-0229-06

习近平总书记深刻指出当今世界正经历百年未有之大变局,新一轮科技革命和新兴产业变革加速演进,全球科技竞争不断迭代升级^[1]。掌握关键核心技术、抢占创新制高点,将是实现进入创新型国家前列发展目标的关键,而构建高效的创新生态则是核心手段。中国早在 2006 年开始布局创新生态系统,实施国家创新型城市试点行动,截至 2020 年底,已有 78 个国家创新型试点城市,创新生态系统的构建和评价逐渐成为各界关注的焦点。

关于城市创新生态系统,Cooke 等^[2]认为是创新主体之间形成的网络模式,基于创新环境通过创新活动产出创新成果,实现协同持续发展。Reynolds 和 Uygun^[3]认为,引入生态概念的区域创新系统能更好诠释系统动态演化过程;靖鲲鹏和宋之杰^[4]认为健康的区域技术创新生态系统利于提升国家创新软实力,并从创新主体规模、资源投入、产出效率、环境适宜 4 个维度构建区域技术创新生态系统健康度指标评价体系。Zemtsov 和 Kotsemir^[5]表示名校和专利资源丰富的地区创新系统效率更高。肖德云、刘俐君和卫武^[6]应用 NCA 和 fsQCA 方法发现,高资产规模构成高创新生态绩效的必要条件。杨西春^[7]基于 DEA-BCC

模型和 DEA-Malmquist 模型,认为创新资源投入是城市创新能力强弱的首要因素,技术进步则是主导要素。

上述研究为科技创新生态系统研究奠定基础,并提供较好借鉴,但相关研究主要以区域或省域为对象,鲜少从城市层面开展评价,且多聚焦于静态视角,较少引入时序维度。而城市是区域科技创新生态的最基本空间单元,同时直辖市、副省级城市是各区域中心城市,在经济发展和科技创新方面均起到新时代排头兵作用,有力支撑了各区域城市群乃至国家的发展,其科技创新经验对于其他城市的科技产业规划和科技创新管理工作均具有重要借鉴意义。因此,本文将应用全局主成分和聚类分析的方法评价直辖市、副省级城市的创新生态系统,量化城市创新差距和短板,为城市创新生态系统优化建言献策,为其他城市的创新发展提供参考。

1 研究设计

1.1 评价指标体系

借鉴前述成果,遵循科学性、代表性、可靠性和数据可获取性原则,从创新环境、创新支付、创新动能、发展趋势、开放水平 5 个维度选取指标,构建重点城市创新生态系统评价指标体系(表 1)。

收稿日期:2022-01-24

作者简介:蔡红(1988—),女,河南信阳人,深圳市维度数据科技股份有限公司,研究副总监,中级经济师,管理科学与工程硕士,研究方向为产业经济、科技管理。

表 1 城市创新生态系统评价指标体系

一级指标	二级指标	指标类型
创新环境	人均 GDPX ₁ /万元	+
	当年出台创新政策措施数量 X ₂ /件	+
	研发投入强度 X ₃ /%	+
	每万人专利授权数 X ₄ /件	+
	每万人专利申请数 X ₅ /件	+
创新支付	地方财政科技支出占地方财政支出比重 X ₆ /%	+
	城镇人均可支配收入 X ₇ /元	+
	人均财政教育经费支出 X ₈ /元	+
	城市每平米住房均价与月平均工资收入比例 X ₉ /(月/m ²)	-
创新动能	国家级高新技术企业数 X ₁₀ /家	+
	高等院校数量 X ₁₁ /所	+
	全球研发 2500 强研发企业中本土企业家数 X ₁₂ /家	+
发展趋势	GDP 增长率(可比价)X ₁₃ /%	+
	固定资产投资增长率 X ₁₄ /%	+
	高新技术企业营业收入增长率 X ₁₅ /%	+
	高新技术企业科技活动人员增长率 X ₁₆ /%	+
开放水平	境内外上市企业数量 X ₁₇ /家	+
	高新技术企业从业人员中留学归国和外籍常驻人员所占比重 X ₁₈ /%	+
	进出口依存度 X ₁₉ /%	+
	实际利用外资额 X ₂₀ /亿美元	+

1.2 研究方法

1.2.1 全局主成分分析

全局主成分分析考虑了指标数据的动态性,能够系统地、动态地、全局地对城市创新生态进行差异性评价^[8]。

1)建立时序立体数据表。评价城市为 n 个,描述指标 m 个,则在 t 年度就有一张 n 行 m 列的数据表: $X^t = (X_{ij})_{n \times m}$ 。 T 年共 T 张数据表,称时序立体数据表。将数据表按时间从上到下排成的矩阵,即为全局数据表:

$$X = (X^1, X^2, \cdots, X^T)_{T \times n \times m} = (X_{ij})_{T \times n \times m} \tag{1}$$

2)原始数据无量纲化,逆向指标:

$$X'_{ij} = \frac{\max(X_j) - X_j}{\max(X_j) - \min(X_j)} \tag{2}$$

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{\delta_j}, i = 1, 2, \cdots, n; j = 1, 2, \cdots, m \tag{3}$$

3)相关系数矩阵:

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{pmatrix} \tag{4}$$

式中, r_{ij} 为第 i 个城市第 j 指标的相关系数, $r_{ij} = r_{ji}$ 。

4)特征值与特征向量。解相关系数矩阵 R 的

特征方程 $|\lambda E - R| = 0$, 得 m 个特征值 ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_i \geq 0$), 及其对应特征向量 $U_1, U_2, \cdots, U_m$ 。其中 $U_j = (\mu_{1j}, \mu_{2j}, \cdots, \mu_{mj})^T$, 由特征向量组成 m 个新的指标变量。

5)确定主成分个数。主成分 Z_i 的贡献率

$$W_i = \lambda_j / \sum_{j=1}^m \lambda_j \tag{5}$$

累计贡献率为

$$\theta_p = \sum_{j=1}^p \lambda_j / \sum_{j=1}^m \lambda_j \tag{6}$$

根据选取原则,特征值大于 1 且累计贡献率达 80%~95% 的特征值数即为主成分个数。

6)计算主成分。

$$F_i = u_{i1} X_1 + u_{i2} X_2 + \cdots + u_{im} X_m, i = 1, 2, \cdots, p \tag{7}$$

7)综合得分:

$$Y = C_1 F_1 + C_2 F_2 + \cdots + C_p F_p \tag{8}$$

式中, $C_i = \lambda_j / \sum_{j=1}^p \lambda_j$ 。

1.2.2 聚类分析

聚类分析是根据在数据中发现的描述对象及其关系的信息,将相似或相关的数据对象划分为一组(簇)。其中系统聚类适用于绝大多数的数据类型,谱系图直观清晰,分类简单有效。

1.3 数据来源

考虑到数据的可靠性和连续性,指标数据主要来源于 2017—2019 年各地统计年鉴、统计公报、政府网公开资料、欧盟全球研发投入 2500 强榜单。

2 实证分析

应用统计软件 SPSS26 对 19 个重点城市 2017—2019 年指标数据进行全局主成分分析与聚类分析,相关统计量结果如下:

2.1 全局主成分分析

2.1.1 适应性检验

基于标准化数据结果,进行 KMO 检验和 Bartlett 球形检验,结果见表 2。根据 Kaiser 因子分析标准, KMO 应大于 0.6^[9], KMO = 0.737; 同时, Bartlett 球形度检验 Sig. = 0.000 < 0.01, 检验通过,认为数据序列适于全局主成分分析。

表 2 KMO 检验和巴特利特检验

KMO		0.737
巴特利特球形度检验	近似卡方	1 226.649
	自由度	190
	显著性	0.000

2.1.2 总方差解释

从表 3 可知,旋转后前 5 个成分的特征值大于

1,累积贡献率达 80.248%,故提取 5 个主成分替代原 20 个指标,进行城市创新生态系统评价。

表 3 总方差解释

成分	初始特征值			提取载荷平方和			旋转载荷平方和		
	总计	方差/%	累积/%	总计	方差/%	累积/%	总计	方差/%	累积/%
1	8.904	44.522	44.522	8.904	44.522	44.522	6.625	33.123	33.123
2	2.903	14.514	59.036	2.903	14.514	59.036	4.409	22.045	55.167
3	1.983	9.916	68.952	1.983	9.916	68.952	1.886	9.432	64.599
4	1.254	6.270	75.221	1.254	6.270	75.221	1.566	7.830	72.429
5	1.005	5.027	80.248	1.005	5.027	80.248	1.564	7.819	80.248
6	0.853	4.263	84.511						
7	0.706	3.531	88.042						
8	0.511	2.557	90.598						
9	0.395	1.976	92.575						
10	0.345	1.725	94.300						
11	0.283	1.417	95.716						
12	0.230	1.148	96.865						
13	0.208	1.039	97.904						
14	0.129	0.645	98.549						
15	0.104	0.518	99.067						
16	0.077	0.384	99.451						
17	0.049	0.244	99.694						
18	0.030	0.151	99.845						
19	0.023	0.113	99.959						
20	0.008	0.041	100.000						

2.1.3 旋转后得分系数矩阵

由表 4 可知, F_1 在 X_4 、 X_5 、 X_1 、 X_6 、 X_9 、 X_8 、 X_7 上成分载荷较高,反映创新环境和创新支付因素。 F_2 在 X_{20} 、 X_{18} 、 X_{12} 、 X_{17} 、 X_{11} 、 X_{10} 、 X_3 上有较高成分载荷,代表开放水平和创新动能。 F_3 与 X_2 、 X_{19} 强正相性,反映创新环境和开放水平因子。 F_4 与

X_{13} 、 X_{14} 、 F_5 与 X_{16} 、 X_{15} 分别显著正相关,从不同层面反映发展趋势因素。

各主成分得分函数:

$$F_i = u_{i1} X_1 + u_{i2} X_2 + \cdots + u_{i20} X_{20},$$
$$i = 1, 2, 3, 4, 5.$$

(9)

限于篇幅,得分系数表不再展示。

表 4 旋转后的主成分载荷矩阵

指标	F_1	F_2	F_3	F_4	F_5	指标	F_1	F_2	F_3	F_4	F_5
X_4	0.954	0.107	0.035	0.150	0.005	X_{17}	0.315	0.751	0.231	-0.023	-0.011
X_5	0.923	0.131	-0.032	0.179	0.085	X_{11}	-0.232	0.750	-0.465	-0.111	-0.028
X_1	0.872	0.260	0.075	0.130	0.118	X_{10}	0.619	0.678	0.140	-0.090	-0.004
X_6	0.868	0.219	-0.142	0.197	0.036	X_3	0.506	0.595	0.228	0.043	0.109
X_9	-0.817	-0.153	-0.382	-0.055	0.135	X_2	-0.105	0.235	0.775	-0.083	-0.173
X_8	0.798	0.144	0.187	0.154	0.099	X_{19}	0.451	0.178	0.752	0.146	0.014
X_7	0.727	0.284	0.101	-0.051	0.008	X_{13}	0.121	0.063	-0.062	0.805	0.126
X_{20}	0.105	0.885	0.078	0.159	-0.033	X_{14}	0.235	-0.146	0.109	0.751	0.067
X_{18}	0.424	0.823	0.198	-0.076	0.022	X_{16}	0.107	-0.098	0.013	-0.014	0.926
X_{12}	0.487	0.771	0.227	-0.125	0.044	X_{15}	-0.011	0.150	-0.220	0.341	0.768

注:旋转方法为凯撒正态化最大方差法。

2.1.4 综合得分及应用结果

$$Y = 0.331\ 23\ F_1 + 0.220\ 45\ F_2 + 0.094\ 32\ F_3 +$$
$$0.078\ 3\ F_4 + 0.078\ 19\ F_5$$

(10)

若 $Y=0$,表明创新生态系统处于 19 个样本城市的平均水平; $Y>0$,表明高于平均水平; $Y<0$,则

低于平均水平^[10]。将 19 个重点城市数据代入,得分见表 5。

从变化趋势看,仅上海、济南、哈尔滨连续 3 年分别维持在第 3、第 14 和第 19 名,其中上海得分持续高位提升。北京、沈阳、南京、杭州、武汉、成都

6 个城市排名有所上升,仅北京、杭州得益于综合得分递增。天津、大连、宁波、青岛综合得分与排名均呈波动变化。重庆、厦门、深圳、长春、广州、西安排名有所下滑,但综合得分基本呈上升趋势。从得分水平看,小于 0 的城市数从 2017 年 13 个降至 2019 年的 9 个,整体呈向好态势;但区域间差距拉大,2019 年第 1 和第 19 名的差距为 2.06,较 2017 年扩大 0.09。从主成分得分看,2019 年深圳、广州、南京 F_1 主成分依次排名前 3,创新环境和创新支付因素领先;北京、上海、武汉为 F_2 主成分得分前 3,开放水平和创新动能指标较优;青岛、上海、厦门 F_3 主成分得分佼佼,创新环境和开放水平因子表现较好;武汉、成都、青岛 F_4 主成分亮眼,长春、济南和北京 F_5 主成分领先,创新趋势指标优势明显。

2.2 聚类分析

基于创新生态系统评价指标,应用系统聚类分析得到图 1 所示谱系图,可将 19 个重点城市创新生态系统分为 4 组:北京、深圳位于第 1 梯队,二者创新生态系统综合得分遥遥领先,其中北京得益于其拥有的占据全国 1/5 的高校资源,而深圳主要由于其浓厚的双创氛围和强大的支付能力。上海、杭州、广州处于第 2 梯队,其综合得分分别位于第 3、第 4 和第 5 名,3 个城市创新支付、创新动能和创新环境指标方面相对处于较优水平。第 3 梯队城市主要包括大连、济南、天津、西安、成都、南京、厦门、宁波、武汉、青岛共

表 5 2017—2019 年中国 19 个城市创新生态系统综合得分及排名

城市	2017		2018		2019	
	得分	排名	得分	排名	得分	排名
北京	0.92	2	0.96	2	1.23	1
天津	-0.24	12	-0.26	15	-0.17	13
上海	0.42	3	0.66	3	0.74	3
重庆	-0.40	15	-0.63	17	-0.53	17
大连	-0.41	16	-0.09	11	-0.41	15
宁波	-0.06	8	-0.11	12	0.05	10
厦门	0.05	6	-0.03	8	0.09	8
青岛	-0.07	9	-0.02	7	0.07	9
深圳	1.12	1	1.30	1	1.14	2
沈阳	-0.71	18	-0.53	16	-0.48	16
长春	-0.51	17	-0.66	18	-0.67	18
哈尔滨	-0.85	19	-0.93	19	-0.83	19
南京	-0.17	11	0.26	6	0.23	6
杭州	0.25	5	0.35	4	0.39	4
济南	-0.29	14	-0.24	14	-0.20	14
武汉	-0.10	10	-0.03	8	0.12	7
广州	0.33	4	0.33	5	0.30	5
成都	-0.26	13	-0.12	13	-0.07	11
西安	-0.04	7	-0.06	10	-0.15	12

10 个城市,各城市均在某个或某些领域有较好表现,如青岛在创新环境、开放水平及发展趋势等方面较为亮眼,武汉的创新动能、创新趋势指标突出。哈尔滨、重庆、沈阳、长春则构成第 4 梯队,其致命弱势在于科技型企业集聚度较低,研发产出水平处于低位,在创新趋势、创新环境及开放水平方面劣势明显。

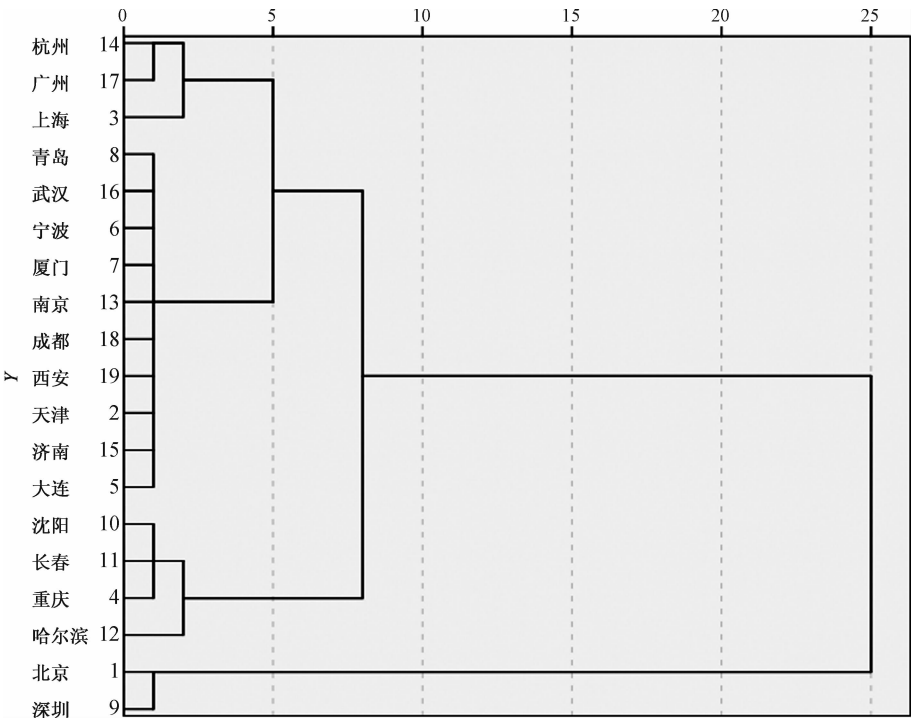


图 1 19 个重点城市创新生态聚类分析图

3 结论与对策建议

3.1 结论

1)创新环境、创新支付影响显著。深圳体现较为明显,其主成分 F_2 、 F_3 、 F_4 、 F_5 在2019年分别排名第11、9、7和16,但 F_1 排名第1,其综合得分仍能保持第2水平。究其原因,创新环境包括经济基础和创新氛围,不仅为创新生态提供有力支撑,也能激发全社会创新活力。

2)创新生态系统的重视程度提高,但得分差距拉大。19个重点城市的创新生态得分整体呈逐年增长态势,2017—2018年,68.4%的城市得分呈现不同程度的增长,其中宁波、上海、重庆等增长速度超过50%。2018—2019年,大连、武汉、等多个城市综合得分增长速度均在1.5~3.6倍水平。2019年第1和第19名的差距为2.06,较2017年扩大0.09。

3)各城市创新生态系统差异明显,各有优势。北京、深圳综合得分为1.23和1.14,远远甩开上海的0.74;但深圳仅在创新环境和创新支付方面优势明显;北京各方面水平相对均衡领先;上海在开放水平、创新环境和创新动能均表现较好,但增长乏力;武汉、厦门、青岛、成都和长春排名较为靠后,但创新环境和创新趋势指标较领先。

3.2 对策建议

1)提升创新“硬+软”环境品质。将优化创新环境、提升环境品质摆在全局位置,一是提升“硬环境”,做大做强经济总量,增强产业支撑,提高经济密度,为创新提供坚实的基础“硬件”。二是优化“软环境”,营造浓厚创新氛围,激活城市创新潜在活力,并将潜在活力转换为实实在在的创新成果。

2)建立完善全生态创新系统。一是健全创新体制机制,作为全生态创新系统的关键环节,能够提供有效指引和方法。二是构建全过程创新生态链,从基础研究、技术攻关、成果产业化、科技金融到人才支撑等,构建全过程全方位的创新生态系统,同时开展找差距补短板行动,针对薄弱环节展开集中攻坚,建立“揭榜挂帅”制度,全力补齐补强创新生态链。

3)培育城市创新强大动能。着力引育国家高新技术企业,鼓励企业加大研发投入、实施关键技术攻关。鼓励支持社会资源参与创新资源项目,培育

孵化优质企业。促进产学研一体化,强化科技成果转化。建设人才集聚区,培育高端科技人才,加强交流。优化服务体系,挖掘科技金融服务潜力。

4)促进跨区域创新系统协同发展。中国经济高质量发展不断推进,区域经济协同化、一体化已成为主趋势,长三角、粤港澳、京津冀、成渝等城市群发展,不仅强调经济协调性,更应强调创新引领,着力探索跨区域优势资源组合。一是根据需要成立政府部门主导的区域创新合作组,搭建高效对接联动机制,实现资本、人才、技术等创新资源的集合集聚;二是建立区域科技创新资源共享平台,实现创新数据共建共治;三是建立区域创新要素自由流动机制,推动高端人才等跨区域服务,提升创新系统的开放度和自由度^[11]。

参考文献

- [1] 吴晓波.超越追赶的新型城市创新生态体系重构[J].杭州,2021(22):28-31.
- [2] COOKE P,GOMEZ URANGA M,ETXEBARRIA G. Regional innovation systems: Institutional organizational dimensions[J]. Research Policy,1997,26(4-5):475-491.
- [3] REYNOLDS E B, UYGUN Y. Strengthening advanced manufacturing innovation ecosystems: The case of Massachusetts [J]. Technological Forecasting and Social Change,2018,136:178-191.
- [4] 靖鲲鹏,宋之杰.健康度视角下区域技术创新生态系统的进化与提升路径:京津冀与长三角的实证研究[J].企业经济,2022(1):143-152.
- [5] ZEMTSOV S,KOTSEMI R. M. An assessment of regional innovation system efficiency in Russia: The application of the DEA approach [J]. Scientometrics, 2019, 120 (2): 1-30.
- [6] 肖德云,刘俐君,卫武.政商环境、二元创新战略耦合对创新生态绩效的影响研究[J].北京邮电大学学报(社会科学版),2021,23(6):79-90.
- [7] 杨西春.珠江-西江经济带城市创新能力测度及评价:基于DEA-Malmquist模型[J].社会科学家,2021(8):91-97.
- [8] 万伦来,干俊峰,余晓钰.基于Matlab的时序全局主成分分析方法及应用[J].华东经济管理,2010,24(1):150-153.
- [9] 李艺.基于因子分析和聚类分析法的安徽省城市技术创新能力评价[J].科技管理研究,2013,33(15):55-59.
- [10] 颜双波.基于主成分分析的县域经济创新能力评价研究:以海上丝绸之路起点泉州为例[J].科技和产业,2016,16(9):16-22.
- [11] 方力,任晓刚.关于推进我国区域科技创新协同发展的思考[J].科技智囊,2020(1):5-8.

Evaluating Research into the Urban Innovation Ecosystem Based on the Methods of
Global Principal Component and Cluster Analysis:

Taking 19 key cities in China as example

CAI Hong

(Shenzhen Dimension Data Technology Co. , Ltd. ,Shenzhen Guangdong 518000,China)

Abstract: Based on the innovation elements of 19 key cities,the evaluation index system of urban innovation ecosystem is built, including five dimensions and twenty secondary indicators,that is innovation environment, innovation payment, innovation kinetic energy, development trend and openness level. Principal components and calculate comprehensive scores are extracted with the method of global principal component analysis, in accordance with the index data from 2017 to 2019,and the urban innovation ecosystem is classified are put forward by utilizing the method of hierarchical cluster analysis. The results show that innovation environment and innovation payment have significant effects on the urban innovation ecosystem. The urban innovation ecosystem have aroused more and more attention of of various cities, but the point difference expand. Significant difference is observed among the cities. Therefore, the following suggestions improving the quality of “hard innovation environment” and “soft innovation environment ”,establishing and optimize the whole ecological innovation system,cultivating strong driving force, and accelerate the coordinated development of cross-regional innovation systems.

Keywords: the urban innovation ecosystem;global principal component analysis;cluster analysis;evaluation;coordinated development