

基于 K-均值聚类与贝叶斯判别的 区域创新极培育能力评价

——以我国 30 个省市自治区为例

汪欢欢

(杭州市社会科学院, 杭州 310026)

〔摘要〕 本文构建了涵盖 4 个一级指标和 12 个二级指标的区域创新极培育能力评价指标体系, 以我国 30 个省市自治区为研究对象, 运用 K-均值聚类分析和贝叶斯判别方法对区域创新极培育能力进行定量评价, 分析不同区域间创新极培育能力存在差异的原因, 提出了提高区域创新极培育能力的政策建议。结果显示: 我国 30 个省市自治区的区域创新极培育能力可以分为 4 类, 从 I 类到 IV 类区域创新极培育能力逐渐降低, 产业创新活力、区域知识支撑水平、区域创新极培育环境、区域经济发展水平逐渐变差, 运用贝叶斯判别能够快速判别区域创新极培育能力大小, 对区域创新极培育的研究具有一定指导意义。

〔关键词〕 创新极 K-均值聚类 贝叶斯判别 培育能力 评价指标体系 区域创新

DOI: 10.3969/j.issn.1004-910X.2019.05.017

〔中图分类号〕 F124.3; F224 〔文献标识码〕 A

引言

区域创新极是指对某一经济区域具有明显创新带动性、支撑性、影响力的创新主体, 在经济全球化和知识经济的背景下, 区域创新极已经成为区域经济社会发展的主要动力源, 面对资源环境的约束, 利用有限资源培育和壮大创新极是促进产业升级、提高国家和区域竞争力最有效的方式^[1,2]。为了更好促进创新要素的自由流动和资源共享, 提升区域创新极对区域经济发展的辐射带动作用, 我国针对培育和壮大区域创新极制定了一系列政策措施, 与此同时, 加大力度提升区域创新极培育能力、尽快形成实现高质量发展和经济社会可持续发展的动力与保障已成为各界关注的焦点。我国幅员辽阔, 由于经济、历史、地域文化和体制机制等多方面原因, 各省市自治区在创新基础和资源能力上存在较大差异, 创新实力、区域创新条件等要素条件不同, 区域创新极培育和发展情况差距明显。系统评估区域创新极的培育能力,

明确不同区域创新极培育所面临问题, 制定与之相匹配的政策制度, 对于提升区域创新极培育能力具有重要实践价值。

目前, 关于区域创新极的研究文献中针对区域创新极培育能力评价研究的成果相对较少。本文以我国 30 个省市自治区为研究对象, 运用 K-均值聚类分析和贝叶斯判别开展区域创新极培育能力评价研究, 以期为进一步培育和壮大区域创新极, 推动区域经济实现高质量发展提供理论支撑。

1 评价方法

1.1 指标体系构建

2002 年, Pierre Mohnen 和 Christian DeBresson 等人在第十四届国际投入产出技术大会上^[3], 提出了“创新极”(Innovative Poles)概念, 指出广东、浙江、上海等是中国的创新极。随后国内外学者李炜、邓向荣、柳卸林等针对区域创新极进行了大量的研究, 这些成果为本研究的开展奠定

收稿日期: 2019-02-21

基金项目: 2016 年度杭州社会科学规划“人才培养计划”专项课题“新产业革命背景下服务业创新的理论演进——兼论与制造业创新的关系”(项目编号: 2016RCZX29)。

作者简介: 汪欢欢, 杭州市社会科学院助理研究员, 博士。研究方向: 区域经济、产业经济。

了基础^[1,3,4,8]。其中针对创新极培育能力的研究归纳起来包括区域创新极培育环境、区域知识支撑水平、区域经济发展水平和产业创新活力4个方面^[5]。

在区域创新极培育环境方面,Porter (2000)^[6]指出区域创新极培育环境和创新基础对创新极的培育能力具有重要影响,投资环境条件、区位优势、政策环境是创新极能够稳定成长和持续发展的保障。韩胜娟 (2010)^[7]认为区域创新极培育需要良好的创新和投资环境,没有良好的创新环境,创新极不可能获得长期的培育绩效。柳卸林 (2003)^[8]认为区域创新环境包括区域创业水平、投资水平、市场需求和科技政策等指标,它是影响区域创新绩效的重要因素,并以我国南方沿海地区为例对研究结果进行了验证。黄鲁成等 (2000)^[9]的研究认为区域创新环境在区域创新极的培育中起到了基础作用,包括创新软环境和创新硬环境两个方面,能够有效提升区域创新投入和产出效率。

在区域知识支撑水平方面,Koch (2006)^[10]指出作为创新主体持续发展与长期生存的重要推动力,协同创新在创新极培育过程中发挥着越来越重要的角色,创新主体与科研机构、中介机构、大学之间的合作为创新极培育提供了更丰富的知识来源。Duysters (2011)^[11]也认为协同创新主体对创新极培育过程中,能够形成相互促进的协同关系,将各自优势发挥到最大,在具有高度复杂性的联盟中形成协同创新效应。张辉^[12]指出创新极培育是多方主体共同作用的结果,区域科技人力资源丰富与否直接影响着创新极的培育效率,科技人力资源是区域知识创新能力的重要组成,为

区域创新极培育提供高效运转的血液。

在产业创新活力方面,马春艳等 (2010)^[13]认为产业技术水平对产业创新具有重要影响,研发基础与能力受到产业发展不平衡性的制约,充足的创新产业投入能够促进创新活动的有效开展。白永亮等 (2013)^[14]建立了以技术创新而非人力资本积累为基础的R&D内生增长模型,通过分析发现R&D活动能够发明新产品或改进技术,并指出国内R&D投入是获取和保持长期竞争优势的必备条件。萧志泳等 (2008)^[15]研究了文化及社会规范与创新和创造的关系,认为随着全球济化水平的提高,文化的熏陶以及经济发展冲击力不断增强,年轻人不甘落后更倾向于创新和创业。

在区域经济发展水平方面,贾明琪 (2015)^[16]认为FDI的扩散及溢出效应对推动受益方技术集中化起着重要作用,外资在促进区域经济发展、产业结构升级以及激励创新等方面发挥着不可替代的作用。李二玲 (2018)^[17]开展了不同尺度区域创新能力与经济发展水平的关系研究,认为区域经济发展水平是区域创新能力提升的重要外部因素,同时区域创新能够反作用于区域经济发展,两者存在互动关系。牛泽东 (2011)^[18]等学者开展了FDI对我国高新技术产业自主创新能力影响的实证分析研究,FDI和产业国际竞争力对创新能力和创新水平起到刺激作用,为创新极培育提供了更多的资源和技术支持。

本文在文献调研的基础上,根据区域创新极培育能力所涉及的基本内容并结合对部分专家的访谈结果,构建了区域创新极培育能力评价指标体系(见表1)。

表1 区域创新极培育能力评价指标体系

一级指标	二级指标
P1 产业创新活力	P11 企业研发经费总额占主营业务收入比重
	P12 企业研发人员数
	P13 百名研发人员拥有发明专利数

续 表

一级指标	二级指标
	P21 高校和科研机构 R&D 经费内部支出中来自企业资金所占比重
P2 区域知识支撑水平	P22 发表科技论文中国外发表论文所占比重
	P23 科研机构 R&D 人员中博士人员所占比重
	P31 技术市场成交额
P3 区域创新极培育环境	P32 各地区高技术产业投资额
	P33 地区财政科技支出占地方财政一般预算支出比重
	P41 人均 GDP
P4 区域经济发展水平	P42 获得外商投资额
	P43 出口额/全国

1.2 评价原理及步骤

1.2.1 K-均值聚类分析

K-均值聚类是一种非监督学习的聚类方法,具有运算速度快、计算量小的特点,广泛应用于分类问题中。假设分类问题有 N 个样本,每个样本有 p 个特征参数,设定聚类个数是 K 。K-均值聚类的计算过程如下:

(1) 随机选择 K 个初始凝聚点作为初始种子节点。

(2) 选择某一样本,计算该样本距离 K 个种子节点的距离,将样本划分到其距离最小的那一类 $C(l)$,当该样本的类别发生改变时,需要对凝聚点重新计算。

$$d(x_i, x_j) = \left[\sum_{r=1}^p |x_{ir} - x_{jr}|^2 \right]^{\frac{1}{2}} \quad (1)$$

$$C(l) = \arg \min_{1 \leq l \leq K} d(x_i, v_l), i = 1, 2, \dots, N \quad (2)$$

$$v_l = \arg \min_v \sum_{i \in C_l} d(x_i, v), i = 1, 2, \dots, N \quad (3)$$

式中, $d(x_i, x_j)$ 代表该样本在聚类时计算的欧式距离; x_i 为第 i 个样本; x_{ir} 为第 i 个样本的第 r 个特征参数; $C(l)$ 为 l 类包含的样本集合; v_l 为 l 类的重心。

(3) 重复上述步骤进行迭代,达到迭代终止条件时终止聚类过程。

1.2.2 贝叶斯判别分析

贝叶斯判别方法是在样本分类信息基础上总结各类别规律,确定待样本类别归属的一种方

法。在进行待样本类别确定时同时考虑了各类别出现的先验概率和错判造成的损失。具体判别过程如下:

(1) 计算各类别先验概率 q_g 和概率密度,在抽取各类样本后建立参数矩阵,设 x_{gij} 为第 g 组第 i 个样本第 j 个参数, \bar{x}_{gj} 为其中参数 j 的平均值, n_g 为第 g 组样本个数。

$$\sum_{g=1}^G q_g = 1 \quad (4)$$

(2) 计算各类别组内协方差矩阵 S 及其逆矩阵 S^{-1} 和行列式, s^{kj} 为 S^{-1} 中 k 行 j 列参数,根据多元正态概率密度函数建立多总体判别函数 $y_g(x)$, C_g 为判别函数系数。

$$s_{kj} = \sum_{g=1}^G \sum_{i=1}^{n_g} (x_{gij} - \bar{x}_{gj})(x_{gik} - \bar{x}_{gk}) / (n_g - G) \quad (5)$$

$$\bar{x}_{gj} = \frac{1}{n_g} \sum_{i=1}^{n_g} x_{gij} \quad (6)$$

$$S = [s_{kj}]_{p \times p} \quad (7)$$

$$S^{-1} = [s^{kj}]_{p \times p} \quad (8)$$

$$y_g(x) = \ln q_g - \frac{1}{2} \sum_{j=1}^p c_{gj} \bar{x}_{gj} + \sum_{j=1}^p c_{gp} x_j \quad (9)$$

$$C_g = \bar{X}_g^T S^{-1} \quad (10)$$

$$c_{gi} = \sum_{k=1}^p s^{kj} \bar{x}_{gk} \quad (11)$$

其中, G 为矩阵 S 的组个数; c_{gi} 为 C_g 中第 g 行第 j 列的元素。

(3) 将样本的各类属性带入到判别函数中,

将该样本划分到概率最大的那一类。

(4) 在检验判别有效性时, 计算该样本被划分到上步中哪一类的后验概率, 并将之前的分组样本带入到判别函数回判。

本文针对区域创新极培育能力的评价主要包括评价指标体系的构建、研究对象数据搜集、K-均值聚类区域创新极培育能力水平分类、贝叶斯判别确定判别公式以及结果分析, 评价流程如图 1。

2 实证分析

数据主要来源于《中国统计年鉴》、《中国科技统计年鉴》和有关省市自治区的公开数据, 本文选取 2017 年的 30 个省市自治区数据进行实证研究(考虑数据的可获得性, 西藏及港、澳、台地区除外)。

2.1 区域创新培育能力水平分类与判别

本文基于 30 个省市自治区的综合得分排名和分析以及研究内容和目的, 将区域创新极培育能力

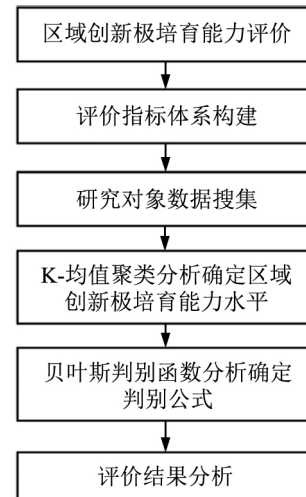


图 1 评价流程图

水平划分为 4 类。经过 K-均值聚类分析确定区域创新极培育能力水平后, 根据已知样本类型, 运用贝叶斯判别分析方法, 建立 4 类区域创新极培育能力水平的 Fisher 线性判别模型, 系数如表 2 所示。

表 2 Fisher 线性判别模型各指标系数

指标	系数			
	I	II	III	IV
P11 企业研发经费总额占主营业务收入比重 (%)	0.633	16.399	-37.706	31.467
P12 企业研发人员数 (万人)	-2.294	0.383	-2.019	0.391
P13 百名研发人员拥有发明专利数 (个)	-0.416	0.383	0.110	0.642
P21 高校和科研机构 RD 经费内部支出中来自企业资金所占比重 (%)	1.963	1.374	1.462	1.380
P22 发表科技论文中国外发表论文所占比重 (%)	0.120	0.533	-0.373	0.077
P23 科研机构 RD 人员中博士人员所占比重 (%)	2.383	0.311	2.836	0.178
P31 技术市场成交额 (亿元)	0.088	0.015	0.078	-0.002
P32 各地区高技术产业投资额 (亿元)	-0.004	0.001	0.008	0.005
P33 地区财政科技支出占地方财政一般预算支出比重 (%)	44.628	14.106	37.743	0.830
P41 人均 GDP (万元)	-8.209	2.201	-8.961	2.671
P42 获得外商投资额 (亿美元)	0.109	0.024	0.099	0.007
P43 出口额/全国 (亿元)	-0.004	-0.003	-0.001	-0.002
常量	-558.398	-121.376	-460.240	-76.179

采用交互验证方法对贝叶斯判别函数的判别效果进行验证, 每次随机抽取 25 个研究样本作为训练集, 将其余 5 个样本作为测试集, 以判断函

数的预测价值, 求出分值并进行分类, 重复上述操作步骤 10 次, 输出的验证结果如表 3 所示。

表 3 交互验证表

分类	交叉预测结果				合计	准确率%
	I	II	III	IV		
I	3	0	0	0	3	100.0
II	0	3	0	0	3	100.0
III	0	1	11	1	13	84.6
IV	0	2	1	28	31	90.3

从交互验证的结果来看，I 和 II 类样本的预测准确率达到了 100.0%，III 类样本的预测准确

率达到了 84.6%，IV 类样本的预测准确率达到 90.3%，整体预测准确率达到 90.0%，表明该方法具有较高的预测价值。

2.2 研究结果分析

通过对 30 个省市自治区样本所包含的 5 个评价指标分布规律的分析，将其分为 I、II、III 和 IV 类，分别含有 2、2、7 和 19 个省份，计算不同区域创新极培育能力对应的评价指标参数分布特征表略。

表 4 不同区域创新极培育能力水平评价参数分布特征

分类	归一化值											
	P11	P12	P13	P21	P22	P23	P31	P32	P33	P41	P42	P43
I	0.63	0.16	0.88	0.91	0.86	0.83	0.60	0.06	0.85	0.99	0.66	0.15
II	0.81	0.98	0.74	0.63	0.73	0.54	0.18	0.73	0.85	0.72	0.94	0.75
III	0.37	0.33	0.41	0.81	0.63	0.57	0.10	0.29	0.52	0.59	0.23	0.14
IV	0.18	0.09	0.38	0.86	0.36	0.40	0.03	0.15	0.21	0.37	0.05	0.02

由表 4 可以直观看到，从 I 类到 IV 类区域创新极培育能力水平逐渐变差，处于同一类的地区在某些方面存在相似的特征或者发展趋势，区域创新极培育能力相近。分类结果分析如下：

I 类：北京和上海，区域创新极培育能力水平最高，归一化后 P13、P21、P22、P23、P31、P33 和 P41 7 项指标在 4 类中排名位居第一，P11、P42 和 P43 3 项指标在 4 类中排名位居第二。作为我国首都的北京，聚集着众多一流的高等院校和具有较强研发能力的科研机构，率先引领建设创新型城市，而上海是长三角城市群的核心城市，正在加快建设具有全球影响力的科技创新中心，聚集着国内外各界精英，拥有中国最大的外贸港口和工业基地。优越的条件和强大的支撑能力促进了高创新能力产业的集聚，带动了区域产业结构升级，使得北京和上海不仅在经济发展上具有先天优势，而且在创新极培育方面的综合能力远超前于其他地区。

II 类：江苏和广东，区域创新极培育能力水

平较高，P11、P12、P32、P42 和 P43 5 项指标在 4 类中排名位居第一，P13、P22、P23、P33 和 P41 5 项指标在 4 类中排名位居第二。江苏充分利用与上海相邻的地理位置优势，企业创新能力和投资环境在科技驱动下不断改善，创新主体的综合创新能力得以提升。广东对外开放较早，创新氛围浓厚，在国外先进创新资源的引导下，制造业和第三产业发展全国领先，成为第一经济强省，在创新园区建设方面起步早，在创新极的培育方面优势明显。但这两个地区在高校和科研机构 RD 经费内部支出中来自企业资金占比相对较低，科技人员素质还不够高，制约了区域创新极培育能力的进一步提升。

III 类：天津、辽宁、浙江、安徽、福建、山东和湖北，区域创新极培育能力水平一般，P12、P23 和 P32 3 项指标在 4 类中排名位居第二，其余指标在 4 类中排名位居第三。这一类区域一般属于沿海城市或者在某一区域具有发展优势的关键区域，但由于在产业创新活力、区域知识支撑

水平、培育环境和经济发展水平等某方面的短板,导致与I和II类区域存在一定差距,阻碍了其区域创新极培育能力水平的提升。如天津是我国北方重要的港口城市和北方第一个自由贸易试验区,曾被誉为中国经济第三增长极,人均GDP曾经连续居于全国前列,但是由于历史的原因,科技中介机构的开发能力还不够高,阻碍了科技成果转化能力,影响了区域创新极的培育。浙江省近年来发展速度较快,紧紧抓住大数据和云计算等智能化发展契机,产业创新活动和区域知识支撑水平均处于较高水平,但技术市场成交额与北京、上海等地区相比,仍处于较低水平。

IV类:其余19个省市自治区,区域创新极培育能力水平较低,除P21和P32 2项指标外其余在4类中排名都在较低水平。这类地区大部分属于内地省份,经济水平、创新极培育环境、产业创新、区域知识支撑水平均处于较低水平,综合创新能力有待加强。随着我国创新驱动发展战略和区域协调发展战略等重大国家战略的深入实施,内地省区和西部地区将迎来新的发展契机,其发展潜力和前景巨大。如河北省面临着京津冀协调发展战略的重大战略机遇,雄安新区的开发建设稳步推进,其目标之一是建设国际一流创新型城市,发展高端高新产业,吸纳和集聚创新要素资源,成为引领区域发展的标杆。四川和重庆经济发展迅速,区位优势不断改善,带动了成渝城市群快速发展。

3 政策建议

3.1 大力发展经济,为区域创新极培育提供保障

北京与上海人均GDP在全国领先,江苏与广东外商投资规模和出口额全国领先,而区域创新极培育能力水平较低的19个省份经济发展水平相对较差,表明区域经济发展水平是影响创新极培育能力的重要因素。良好的经济基础能够为区域创新极培育提供有力支撑,促进产业创新活力提升、实现科技人才流动和配置,有效承接科技成果转移转化。经济发展水平越高,区域创新溢出

能力和扩散效果越好,因此,为了进一步提升区域创新极培育能力,应加大力度发展地方经济,推动经济实现高质量发展。

3.2 加大研发投入,提高创新产出效率

提高研发投入强度,鼓励企业开展创新活动,促进创新成果转移转化,为产业发展提供创新支持,提升产业创新活力,为区域创新极的培育和发展提供载体和支撑。对于能够明显带动区域发展的产业领域,加大财政科技投入力度,鼓励开展大型科技基础设施建设和产业关键技术攻关,提高企业自主创新能力,推动区域创新极的形成。

3.3 重视协同创新,发挥创新主体能力

广泛吸纳高尖端科技人才,不断提升科技人员创新能力和水平是提高创新极培育能力的突破点,加大产学研协同创新合作,推动区域创新网络形成,为区域创新极的形成和壮大提供支撑。科研机构能够创造出大量的科研成果,是重要的科研主体,对区域协同创新和区域创新极培育水平具有较大的影响,应高度重视独立科研机构的开发能力和新型科研机构的培育,提升科技中介服务机构的服务能力,为区域创新极的培育提供载体平台。

参 考 文 献

- [1] 李炜,李子彪,康凯. 区域创新极培育能力综合评价研究[J]. 技术经济与管理研究, 2018, (4): 49~53.
- [2] 董微微. 政府干预对创新集群演化过程的影响研究——基于BBV模型的仿真[J]. 科技管理研究, 2016, 36(18): 7~12.
- [3] 邓向荣,周密,李伟. 我国科技创新极化度指数的构造及区域比较[J]. 财经研究, 2007, (6): 67~76.
- [4] 李炜,李子彪,康凯. 区域创新极培育能力影响因素实证研究[J]. 技术经济与管理研究, 2018, (6): 117~122.
- [5] 董微微,蔡玉胜. 我国国家自主创新示范区创新能力评价[J]. 工业技术经济, 2018, 37(8): 78~85.
- [6] Porter M E. Clusters and Competition. New Agenda for Companies [J]. On Competition, 1998, (3): 197~287.
- [7] 韩胜娟. 基于因子分析的区域投资环境评价[J]. 企业经济, 2010, (8): 24~26.
- [8] 柳卸林. 区域创新体系成立的条件和建设的关键因素[J].

- 中国科技论坛, 2003, (1): 18~22.
- [9] 黄鲁成. 关于区域创新系统研究内容的探讨 [J]. 科研管理, 2000, (2): 43~48.
- [10] Koch A, Stahlecker T. Regional Innovation Systems and the Foundation of Knowledge Intensive Business Services. A Comparative Study in Bremen, Munich, and Stuttgart, Germany [J]. European Planning Studies, 2006, 14 (2): 123~146.
- [11] Duysters G, Lokshin B. Determinants of Alliance Portfolio Complexity and Its Effect on Innovative Performance of Companies [J]. Journal of Product Innovation Management, 2011, 28 (4): 570~585.
- [12] 张辉, 石琳. 人力资本与区域创新研究——基于空间面板模型的分析 [J]. 湖南大学学报(社会科学版), 2018, 32 (5): 49~57.
- [13] 马春艳, 马强. 我国农业生物产业技术创新能力评价研究 [J]. 生态经济, 2010, (1): 89~91.
- [14] 白永亮, 王瑶. FDI、研发投入与高新技术产业创新绩效——基于武汉东湖国家自主创新示范区的经验分析 [J]. 特区经济, 2013, (7): 141~144.
- [15] 萧志泳, 林嵩. 澳门创业环境研究 [J]. 中国软科学, 2008, (7): 77~82.
- [16] 贾明琪, 刘双双, 辛江龙. 外商直接投资与科技创新、经济增长——基于西部10省面板数据的实证分析 [J]. 科学决策, 2015, (4): 21~34.
- [17] 李二玲, 崔之珍. 中国区域创新能力与经济发展水平的耦合协调分析 [J]. 地理科学, 2018, 38 (9): 1412~1421.
- [18] 牛泽东, 张倩肖, 王文. FDI对我国高新技术产业自主创新影响力影响的实证分析 [J]. 科技进步与对策, 2011, 28 (18): 51~55.

Evaluation of Regional Innovation Pole Cultivation Ability Based on K-means Clustering and Bayesian Discrimination ——Taking 30 Provinces, Municipalities and Autonomous Regions as Examples

Wang Huanhuan

(Hangzhou Academy of Social Sciences, Hangzhou 310026, China)

(Abstract) The evaluation index system of regional innovation pole cultivation ability covering 4 first-level indicators and 12 second-level indicators is constructed in this paper. Taking 30 provinces, municipalities and autonomous regions in China as the research object, the quantitative evaluation of regional innovation pole cultivation ability level is carried out by using K-means clustering analysis and Bayesian discriminant analysis. The reasons for the differences of regional innovation pole cultivation ability are analyzed, and the policies for improving regional innovation pole cultivation ability are put forward. The results show that the cultivation ability of regional innovation poles in 30 provinces, municipalities and autonomous regions in China can be divided into three categories. The cultivation ability of regional innovation poles from type I to type IV gradually decreases, and the vitality of industrial innovation, the level of regional knowledge support, the cultivation environment of regional innovation poles and the level of regional economic development gradually deteriorate. Bayesian discriminant can be used to quickly identify the cultivation ability of regional innovation poles, which has a certain guiding significance for the research of regional innovation pole cultivation.

(Key words) innovation pole; K-means clustering; Bayes discrimination; cultivate ability; evaluation index system; regional innovation

(责任编辑: 张舒逸)