

我国城市建设水平研究*

陈瑜鑫, 刘惠篮**

(贵州大学 数学与统计学院, 贵阳 550025)

摘要:基于《中国统计年鉴》最新数据,运用稀疏主成分方法及聚类分析方法,对 2017 年和 2018 年我国各地区城市建设水平的差异进行了研究;稀疏主成分分析在降维的同时,通过惩罚函数方法,对载荷矩阵进行稀疏化处理,其分析结果相较于主成分分析具有较强的可解释性;此外,对我国各地区进行了聚类分析,聚类结果验证了稀疏主成分法排名的合理性,结果表明:我国各地区城市建设水平是参差不齐的,江苏、浙江、广东、山东的城市建设水平较高,而西藏、宁夏、贵州的城市建设水平较低;最后,为我国各地区城市化的发展提供了一些对应的建议。

关键词:稀疏主成分;聚类分析;城市建设;设施水平

中图分类号: O212.1 **文献标志码:** A **文章编号:** 1672-058X(2021)04-0111-07

0 引言

由于社会和自然因素影响,我国各地区的城市建设情况及设施水平发展不均衡。本次研究的主要目的是对城市建设情况及设施水平的相关数据进行分析,从而得出各地区城市建设水平的综合得分,更好地了解各地区城市建设情况与设施水平的差异。

近年来,众多研究人员对城市设施水平进行了比较研究,得出了一些研究成果。2004 年,李凤琴^[1]选取了与城市设施水平有关的 7 个变量,在因子分析的基础上,结合聚类分析对各地区城市设施水平进行评价,由聚类分析的结果将这些地区分成 3 类,最终得到了城市设施水平的大致情况;2008 年,张哲等^[2]针对全国 31 个城市基础设施建设数据,运用聚类分析和主成分分析进行实证研究,结论表明:基础设施建设水平较高的是经济发达的北京、上海等城市,相对于南方城市而言,北方城市的建设水平普遍较低;2013 年,谢英欣等^[3]基于 K-均值法

对我国各地区城市基础设施建设水平进行聚类,分析我国东部和西部地区城市基础设施建设不均衡的具体因素,提出政府要充分发挥宏观调控的作用,加大对我国中西部地区城市的政策帮助和资金投入,加快改善中西部落后省份城市的基础设施建设水平;2015 年,张伟芳等^[4]选取了甘肃省各城市 10 年的基础设施数据,运用变异系数法计算各州市的城市基础设施得分,通过图表描绘了各城市基础设施的改变,得出了甘肃省 14 个州市基础设施建设水平的发展存在明显差异,从 3 个方面分析了城市基础设施的影响因素;2016 年,荣月静等^[5]选取长江三角洲地区 16 个城市基础设施建设的数据,基于多因素综合评价等方法对该地区的城市基础设施建设水平进行研究,提出城市基础设施的各个方面要协调发展,同时也指明各城市在基础设施发展过程中存在的问题,并提出与之相对应的解决方法,期望长江三角洲各个城市的基础设施建设为其他城市提供一定的经验;2017 年,朱怡然^[6]选择了城市桥梁、绿地面积、用水普及率等 22 个指标对全国 31 个省市自

收稿日期:2020-07-15;修回日期:2020-09-10.

* 基金项目:国家自然科学基金项目资助(11761020);黔科合平台人才[2017]5788;黔科合重大专项[2018]3001.

作者简介:陈瑜鑫(1997—),男,贵州镇远人,硕士研究生,从事复杂数据研究.

** 通讯作者:刘惠篮(1988—),女,贵州贵阳人,副教授,博士,从事复杂数据研究. Email:hlliu5@gzu.edu.cn.

治区的城市竞争力进行评价,利用主成分分析选取了前 4 个主成分并得到 31 个省市自治区的综合得分,根据综合得分将这些地区进行排序,对城市建设水平较低的地区给出了一些很有针对性的建议。

对于城市建设水平的研究,上述学者建立了比较完善的评价指标体系,为研究最新的城市建设水平提供了坚实的基础。采用不同的多元统计方法研究各地区城市建设及设施水平,得到了不同地区城市建设水平发展的具体情况,针对各地区城市建设水平的差异,提出了与之相对应的解决对策。但是以上研究大多基于主成分分析和因子分析等传统多元统计方法,在某些情形下难以解释。稀疏主成分方法极大地稀疏了载荷矩阵,因此,它相较于主成分方法具有更强的可解释性。通过平衡主成分的稀疏度和解释程度,讨论了惩罚参数的选择方法。

本文基于稀疏主成分法与聚类方法对 2017 年和 2018 年城市建设情况及设施水平的数据进行研究,深入了解各地区城市建设情况与设施水平的差异,得到了 2017 年到 2018 年城市建设水平的动态变化,这些有意义的结论将促进城市的建设与发展。

1 建立模型

1.1 主成分

主成分分析^[7]是在保留绝大多数信息量的条件下,通过线性变换的方式,用较少的新变量替代原来较多的变量。它的实质就是通过原始变量的线性组合来达到降低原始变量维度的目的,主成分就是经过变换后得到的新变量。主成分与原始变量之间存在以下的关系:各主成分之间互不相关,原始变量的数量大于主成分的数量,多个原始变量的不同线性组合构成了不同的主成分,选取的主成分只损失了原始变量很少部分信息。

设 $\mathbf{X}=(X_1, X_2, \dots, X_p)^T$ 是 p 维随机向量,均值 $E(\mathbf{X})=\boldsymbol{\mu}$, 协方差阵 $\text{Cov}(\mathbf{X}, \mathbf{X})=\boldsymbol{\Sigma}$ 。考虑以下的线性变换:

$$\begin{cases} Z_1 = a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p \\ Z_2 = a_{12}X_1 + a_{22}X_2 + \dots + a_{p2}X_p \\ \dots \\ Z_p = a_{1p}X_1 + a_{2p}X_2 + \dots + a_{pp}X_p \end{cases}$$

容易看到:

$$\begin{cases} \text{Var}(Z_i) = \mathbf{a}_i^T \boldsymbol{\Sigma} \mathbf{a}_i, (i = 1, 2, \dots, p) \\ \text{Cov}(Z_i, Z_j) = \mathbf{a}_i^T \boldsymbol{\Sigma} \mathbf{a}_j, (i, j = 1, 2, \dots, p) \end{cases}$$

为了得到较好的结果,希望 Z_i 的方差比较大,同时 Z_i 之间是互相独立的,由于

$$\text{Var}(Z_i) = \mathbf{a}_i^T \boldsymbol{\Sigma} \mathbf{a}_i$$

对于任意的常数 c ,有

$$\text{Var}(cZ_i) = c^2 \mathbf{a}_i^T \boldsymbol{\Sigma} \mathbf{a}_i$$

从上面的式子可以看出,对 \mathbf{a}_i 不加限制时, $\text{Var}(Z_i)$ 将会变得任意大,这样,所研究的问题将变得没有意义。因此,有下面的约定:

- (1) $\mathbf{a}_i^T \mathbf{a}_i = 1$ 。
- (2) Z_i 与 Z_j 相互无关, $i \neq j, i, j = 1, 2, \dots, p$ 。
- (3) $\text{Var}(Z_i) = \max_{\mathbf{a}^T \mathbf{a} = 1, \mathbf{a}_j^T \mathbf{a}_j = 0} \text{Var}(\mathbf{a}^T \mathbf{X})$ 。

则由 \mathbf{a}_1^T 确定的 Z_1 称为 \mathbf{X} 的第一个主成分,由 \mathbf{a}_2^T 确定的 Z_2 称为 \mathbf{X} 的第二个主成分, ..., 由 \mathbf{a}_p^T 确定的 Z_p 称为 \mathbf{X} 的第 p 个主成分。

由上面提及的知识可知,求第一主成分 $Z_1 = \mathbf{a}_1^T \mathbf{X}$, 是求 $\mathbf{a}_1 = (a_{11}, a_{21}, \dots, a_{p1})^T$, 使得在 $\mathbf{a}_1^T \mathbf{a}_1 = 1$ 的条件下, Z_1 的方差达到最大。显然,这是一个条件极值的求解问题。令

$$\varphi(\mathbf{a}_1, \lambda) = \mathbf{a}_1^T \boldsymbol{\Sigma} \mathbf{a}_1 - \lambda(\mathbf{a}_1^T \mathbf{a}_1 - 1)$$

等式两边分别对 \mathbf{a}_1 和 λ 求导,并令其等于 0

$$\begin{cases} \frac{\partial \varphi}{\partial \mathbf{a}_1} = 2(\boldsymbol{\Sigma} - \lambda \mathbf{I}) \mathbf{a}_1 = 0 \\ \frac{\partial \varphi}{\partial \lambda} = \mathbf{a}_1^T \mathbf{a}_1 - 1 = 0 \end{cases}$$

由 $\mathbf{a}_1^T \mathbf{a}_1 = 1$ 可知 $\mathbf{a}_1 \neq 0$, 对于 $(\boldsymbol{\Sigma} - \lambda \mathbf{I}) \mathbf{a}_1 = 0$, 由高等代数的相关理论可知,求上面的方程组实质上就是求解 $\boldsymbol{\Sigma}$ 的特征值和相对应的单位特征向量的问题。一般地,可以通过求 $\boldsymbol{\Sigma}$ 的第 i 大特征值所对应的单位特征向量求 \mathbf{X} 的第 i 主成分。

由于 $\text{Cov}(\mathbf{Z}, \mathbf{Z}) = \text{diag}(\lambda_1, \dots, \lambda_p)$, 记 $\boldsymbol{\Sigma} = (\sigma_{ij})_{p \times p}$, 可知 $\sum_{i=1}^p \lambda_i = \sum_{i=1}^p \sigma_{ii}$ 。因此,第 k 个主成分的方差贡献率是 $\alpha_k = \frac{\lambda_k}{\lambda_1 + \dots + \lambda_p}$, 主成分 Z_1, Z_2, \dots ,

Z_m 的累积方差贡献率是 $\sum_{i=1}^m \lambda_i / \sum_{i=1}^p \lambda_i$ 。从上面可以知道,彼此不相关的主成分的方差之和等于随机向量 \mathbf{X}

的总方差,且存在 $m(m < p)$,使得 $\sum_{i=1}^m \lambda_i \approx \sum_{i=1}^p \sigma_{ii}$, 可以利用前面 m 个主成分替换 p 个原始变量所提供的总方差。减少变量的个数是进行主成分分析的目的之一,分析具体问题时,只选取前面 m 个主成分,为了确定 m 的取值,一般认为取到的 m 使得累积方差贡献率不小于 85% 比较合适,即

$$\frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^p \lambda_i} \geq 85\%$$

另外,主成分的选取个数可以由特征值的变化情况来确定,可以利用 R 软件画出碎石图,通过图形的变化趋势来确定主成分的个数 m 。由相关阵 R 求主成分的过程与由协方差阵 Σ 求主成分的过程是一致的,这是因为 X 的相关阵刚好是 X^* 的协方差阵, X^* 是原始变量经过标准化处理后的变量。

1.2 稀疏主成分

Zou 等^[8]提出了稀疏主成分分析方法,稀疏主成分是建立在主成分可以写成一个二次惩罚的回归型优化问题的基础上的,Lasso^[9]惩罚和 Elastic Net^[10]可以直接整合到回归准则中,从而实现模型具有稀疏加载的主成分。稀疏加载后的主成分,使得载荷矩阵的一部分变量系数压缩为零,不仅提高了模型的可解释性,而且稀疏主成分能正确地识别重要的变量,不会丢失重要变量的信息。通常情况下,稀疏主成分的求解能转化为如下的优化问题:

$$(\hat{A}, \hat{B}) = \arg \min_{A, B} \sum_{i=1}^n \|X_i - AB^T X_i\|^2 + \eta \sum_{j=1}^k \|\beta_j\|^2 + \sum_{j=1}^k \eta_{1,j} \|\beta_j\|_1$$

s. t. $A^T A = I_{k \times k}$

其中: X_i 表示原始矩阵的第 i 列。下面的算法总结了稀疏主成分的求解步骤:

Step 1 令 A 初始化为前 k 个普通主成分的载荷向量 $V[:, 1:k]$ 。

Step 2 对于固定的 $A = [\alpha_1, \dots, \alpha_k]$,其中 $\alpha_i, i=1, 2, \dots, k$ 表示前 k 个主成分的载荷向量,解下面的 Elastic Net 问题:

$$\beta_j = \arg \min_{\beta} (\alpha_j - \beta)^T X^T X (\alpha_j - \beta) + \eta \|\beta\|^2 + \eta_{1,j} \|\beta\|_1$$

其中 $j=1, 2, \dots, k$ 。

Step 3 对于固定的 $B = [\beta_1, \beta_2, \dots, \beta_k]$,其中, $\beta_i, i=1, 2, \dots, k$ 表示前 k 个稀疏主成分的载荷矩阵。计算 SVD 分解 $X^T X B = U D V^T$,然后更新 $A = U V^T$ 。

Step 4 重复第二步至第三步,直到收敛为止。

Step 5 标准化: $\hat{V}_j = \frac{\beta_j}{\|\beta_j\|}, j=1, 2, \dots, k$ 。

通常可以使用 LARS-EN 算法求解步骤 2 中的 Elastic Net 问题,从而得到合适的惩罚系数。同时 LARS-EN 算法很大程度上提高了稀疏主成分的计算效率。

准确地选择惩罚系数对求解主成分是极其关键的,当 $n > p$ 时, η 通常被选取为零,不同的 $\eta_{1,j}$ 可以得到不同的稀疏主成分。本文选择惩罚系数时,不仅要考虑主成分的稀疏度,同时也必须兼顾主成分对于原始变量的解释程度^[11]。令 M 表示载荷为零的变量个数, N 代表总变量个数, PEV 表示方差解释程度。定义 S 为

$$S = \frac{M}{N} \times \text{PEV}$$

用解释方差度刻画解释程度,本文考虑 S 达到最大时所对应的 $\eta_{1,j}$ 值是最优的惩罚系数。

1.3 聚类分析

作为一种常用的多元统计方法,聚类分析将样品进行分类,根据样品之间的相似性,把相似度越高,差异越小的样品聚为一类,使得同一类中样品相似度高,不同类之间样品相似度低。这里主要研究的是 Q 型聚类分析。研究样品之间的关系是对样品进行聚类的前提,有许多统计量可以用来描述样品间的相似程度,一般用得比较多的统计量就是距离。样品 X_i 与 X_j 之间的距离用 d_{ij} 来表示,一般有下面的要求:

$$(1) d_{ij} \geq 0, \forall i, j; d_{ij} = 0 \Leftrightarrow X_i = X_j。$$

$$(2) d_{ij} = d_{ji}, \forall i, j。$$

$$(3) d_{ij} \leq d_{ik} + d_{kj}, \forall i, j, k。$$

此处使用的样品间的距离是欧式距离。其定义为

$$d_{ij} = \left[\sum_{k=1}^p |x_{ik} - x_{jk}|^2 \right]^{1/2}$$

在众多聚类方法中,应用最为广泛的的就是系统聚类法,系统聚类法的聚类过程可以由谱系图最终表示出来。系统聚类法的步骤如下:

Step 1 计算 n 个样品两两之间的距离。

Step 2 将 n 个样品对应于 n 个类,保证每个类只包括一个样品。

Step 3 将 n 个类中距离最接近的两个类合成一个新的类。

Step 4 计算已有的各类与最新得到的类之间的距离;观察类的个数是否为 1。如果不是,重复步骤 3 和 4;若是,则转到步骤 5。

Step 5 画出谱系聚类图。

Step 6 决定分类的个数和聚类的成员。

2 实证分析

本文使用的数据为目前所有的最新数据,来源于 2018 年和 2019 年中国统计年鉴,各地区城市建设情况及设施水平的统计指标见表 1。首先采用稀疏主成分方法对 2018 年各地区的城市建设情况及设施水平进行评价,利用该方法计算各地区的综合得分,按综合得分的高低进行排序。仅选取一年的城市建设水平数据进行分析,只能了解这些地区该年度的城市建设情况,无法更好地了解城市发展的趋势。因此,将 2017 年和 2018 年我国各地区的综合得分进行对比,得出两年间我国城市建设情况及设施水平的变化趋势。同时运用系统聚类法对全国 31 个省市自治区分类。

表 1 城市建设水平评价指标

总目标	评价指标	符号
城市建设 情况	城区面积(km ²)	X ₁
	建成区面积(km ²)	X ₂
	城市建设用地面积(km ²)	X ₃
	征用土地面积(km ²)	X ₄
	城市人口密度(人/km ²)	X ₅
城市设施 水平	城市用水普及率(%)	X ₆
	城市燃气普及率(%)	X ₇
	每万人拥有公共交通工具(标台)	X ₈
	人均城市道路面积(m ²)	X ₉
	人均公园绿地面积(m ²)	X ₁₀
	每万人拥有公共厕所(座)	X ₁₁

由于各个评价指标的单位不一致,而且指标的量级差距较大,所以本文在使用主成分分析(表 2)和稀疏主成分分析之前,需要对数据进行标准化处理。通过稀疏主成分分析法,可以得到前 5 个稀疏主成分的载荷矩阵、贡献率、累积贡献率,惩罚系数 $\eta_{1,j}=(0.3,0.6,0.2,1,0.6)$ 对应于不同的稀疏主成分。稀疏化后的载荷矩阵如表 3 所示。

表 2 主成分的载荷矩阵

变 量	PC1	PC2	PC3	PC4	PC5
X ₁	0.41	0.13	0.22	0.23	0.28
X ₂	0.42	0.08	0.31	-0.19	0.06
X ₃	0.42	0.05	0.33	-0.17	0.07
X ₄	0.34	0.14	0.16	-0.21	-0.36
X ₅	-0.09	-0.50	0.18	-0.63	-0.05
X ₆	0.27	-0.45	-0.37	0.05	0.02
X ₇	0.30	-0.48	-0.26	0.11	-0.10
X ₈	0.25	-0.09	-0.27	-0.07	0.60
X ₉	0.21	0.25	-0.42	-0.31	-0.47
X ₁₀	0.26	0.29	-0.37	0.13	-0.10
X ₁₁	-0.12	0.34	-0.31	-0.55	0.42
贡献率/%	39.5	15.2	14.2	9.3	7.7
累积贡献率/%	39.5	54.7	68.9	78.2	85.9

表 3 稀疏主成分的载荷矩阵

变 量	SPC1	SPC2	SPC3	SPC4	SPC5
X ₁	-0.34	0	0	0	0
X ₂	-0.68	0	0	0	0
X ₃	-0.56	0	0	0	0
X ₄	-0.33	0	0	0	0
X ₅	0	0	-0.65	0	0
X ₆	0	-0.72	0	0	0
X ₇	0	-0.69	0	0	0
X ₈	0	0	0	0	-1
X ₉	0	0	0.47	0	0
X ₁₀	0	0	0.60	0	0
X ₁₁	0	0	0	-1	0
贡献率/%	28.3	15.0	14.0	7.9	6.8
累积贡献率/%	28.3	43.3	57.3	65.2	72.0

城区面积、建成区面积、城市建设用地面积、征用土地面积的综合就是第一主成分,第一主成分主要代表城市建设中的面积;第二主成分为水气主成分,主要与城市用水普及率和城市燃气普及率相关;第三主成分为人均面积及密度主成分,主要表示的是人均城市道路面积、人均公园绿地面积和城市人口密度;第四主成分和第五主成分中,由于其他变量的系数被压缩为零,分别只剩下一个变量的系数非零,因此,第四主成分表示每万人拥有公共厕所;第五主成分表示每万人拥有公共交通工具。

由主成分和稀疏主成分载荷矩阵的对比分析来看,使用稀疏主成分极大地稀疏了载荷矩阵,使得每个主成分所对应的部分变量系数被压缩为 0。由表 3 可知,前 5 个稀疏主成分的方差贡献率逐渐减少,

保留了原始变量 72.0% 的信息,同时极大地稀疏了载荷变量的个数。第一主成分中,载荷为 0 的变量个数有 7 个,第二主成分中,载荷为 0 的变量个数有 9 个,第三主成分中,载荷为 0 的变量个数有 8 个,第四主成分和第五主成分中,载荷为 0 的变量个数都有 10 个。利用稀疏主成分对各地区城市建设情况及设施水平进行分析,可以得到各地区的综合得分及排名如表 4 所示。值得注意的是,表 4 中综合得分负得越大,综合得分越高。

表 4 各地区城市建设水平综合得分及排名

Table 4 Comprehensive score and ranking of urban construction level in various regions

地 区	2018 年		2017 年	
	综合得分	排名	综合得分	排名
北京	-1.003	6	-1.351	3
天津	-0.246	15	-0.860	6
河北	-0.609	9	-0.168	16
山西	0.536	22	-0.218	15
内蒙古	0.104	19	2.635	30
辽宁	-0.034	18	-0.953	5
吉林	0.997	27	0.696	26
黑龙江	-0.966	7	-0.800	7
上海	-0.522	11	-1.907	2
江苏	-2.088	1	-0.282	13
浙江	-1.393	2	-0.508	11
安徽	-0.156	16	0.200	19
福建	-0.534	10	-0.079	17
江西	0.545	23	-0.445	12
山东	-1.177	4	-0.967	4
河南	-1.006	5	-0.679	8
湖北	-0.259	14	-0.672	9
湖南	-0.329	13	-0.672	10
广东	-1.294	3	-2.068	1
广西	1.054	28	0.298	21
海南	0.768	24	0.551	25
重庆	0.925	26	0.480	23
四川	-0.499	12	-0.243	14
贵州	1.331	29	0.740	27
云南	-0.154	17	1.206	28
西藏	3.654	31	2.927	31
陕西	-0.718	8	0.550	24
甘肃	0.770	25	0.130	18
青海	0.131	20	0.394	22
宁夏	1.867	30	1.772	29
新疆	0.302	21	0.292	20

从表 4 可以看到 2018 年我国各地区城市建设

与设施水平的大致情况,具体表现在以下几个方面:综合得分排名靠前的 10 个地区分别是江苏、浙江、广东、山东、河南、北京、黑龙江、陕西、河北、福建,表明这 10 个地区的城市建设情况及设施水平较高。综合得分排名第 11 至 21 名的地区分别为上海、四川、湖南等 11 个地区,表明这些地区的城市建设情况及设施在国内处于中等水平,其他 10 个地区的综合得分排名是比较靠后的,这 10 个地区的城市建设情况及设施水平与全国其他各地区存在一定的差距。

从 2017 年与 2018 年的各地区城市建设情况及设施水平的对比分析来看,综合得分排名呈上升趋势的地区有 12 个,分别是河北、内蒙古、江苏、浙江、安徽、福建、河南、海南、四川、云南、陕西、青海。这 12 个地区中,陕西、江苏、内蒙古和云南的上升幅度是非常明显的,陕西的综合得分排名从 2017 年的第 24 名上升到了第 8 名,江苏次之,从 2017 年的第 13 名上升到了第 1 名,内蒙古的综合得分排名从第 30 名上升到了第 19 名,云南的综合得分排名也上升到了第 17 名。这 4 个地区的城区面积、建成区面积都有了较大幅度的增长,导致了它们的综合排名有了较为明显的提升。这两年间,综合得分排名保持不变的地区共有 3 个,依次是山东、黑龙江和西藏。由表 3 可以看出,山东和黑龙江的综合得分在 2017 年和 2018 年都是比较靠前的,其中,山东的综合得分排名为第 4 名,黑龙江的综合得分排名为第 7 名,西藏的综合得分最低,它的综合得分排名为第 31 名。综合得分排名呈下降趋势的地区有 16 个,分别是广东、上海、北京、辽宁、天津、湖北、湖南、江西、山西、甘肃、新疆、广西、重庆、贵州、吉林、宁夏。这 16 个地区中,辽宁、江西的下降幅度比较明显,辽宁的综合得分排名从 2017 年的第 5 名降到了 2018 年的第 18 名,江西的综合得分排名从 2017 年的第 12 名下降到了 2018 年的第 23 名。

由于各变量的量纲不一致,所以使用标准化后的数据进行分析,计算关于欧式距离的距离矩阵,得到各个样品之间的距离。对 2018 年各地区城市建设水平的相关数据选择离差平方和法进行系统聚类,可以将 31 个地区分为 3 类,聚类结果见图 1。第一类包括广东、江苏、山东;第二类包括天津、上海、云南、陕西、黑龙江、河南、湖南、新疆、河北、福建、海南、青海、江西、山西、贵州、甘肃、内蒙古、宁夏、北京、浙江、安徽、四川、辽宁、湖北、广西、吉林、重庆;第三类包括西藏。

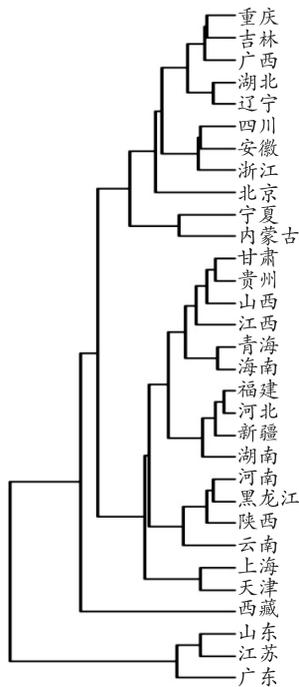


图 1 系统聚类图

Fig. 1 Hierarchical clustering diagram

运用稀疏主成分法与系统聚类法对 2018 年城市建设水平进行综合评价,第一类的 3 个地区的综合得分排名都比较靠前,第二类地区的综合得分排名居中,第三类的西藏地区综合得分排名最后。这也说明了由稀疏主成分得到的综合得分排名是合理的。

3 结论与建议

从 2018 年各地区综合得分的排名来看,我国各地区城市建设及设施水平的发展是存在差异的,总体趋势是东部沿海地区的水平普遍较高,尤其是江苏、浙江、广东、山东的城市建设水平稳居前列;西南和西北内陆地区的城市建设及设施水平较低,特别是西藏、宁夏、贵州与其他地区的水平存在较大的差距。为了保证各地区城市建设水平更加均衡地发展,更好地推进我国城市建设的进程,在此,提出如下几点建议:

综合得分排名名列前茅的 4 个地区分别是江苏、浙江、广东、山东,从它们所处的地理位置来看,这几个省都处于我国东部沿海地区,它们的城市建设及设施水平在全国处于领先。东部沿海地区城市建设水平迅速崛起的重要原因在于它们天然的地理优势,改革开放初期,首先发展沿海的策略使得这些地区的城市化水平较高,所以,东部沿海地区要保持

现在的发展势头,继续巩固现存的发展优势,树立可持续发展理念,实现城市建设又好又快发展;同时要给其他地区的城市建设树立典范,为全国各地区的城市建设提供宝贵的经验。

综合得分排名处于中间位置的这些地区,它们的城市建设及设施水平在全国范围内处于中等。这些地区要向城市建设较好的东部沿海地区看齐,争取与东部沿海地区的城市建设水平接近,让自身的城市建设迈上一个新的台阶,实现的跨越式发展。其中,上海和四川的城市建设水平发展极其显著,其他的地区要积极借鉴这两个地区的城市建设经验,不断提高城市建设水平。

在错综复杂的因素影响下,西南地区 and 西北地区的城市建设水平是较低的,尤其是西藏、宁夏、贵州的城市建设长期处于较低水平。这几个地区位于我国的内陆地区,不仅自身经济实力相对偏弱,而且得到政府的资金帮助有限。因此,西南、西北地区要根据自身优势,因地制宜地发展经济,从而保证城市建设有源源不断的资金支持。同时要充分认识到与其他地区城市建设与设施水平发展的差距,从其他地区吸取经验,摆脱城市建设情况较差的面貌,实现我国各地区城市建设的均衡发展。

参考文献 (References):

- [1] 李凤琴. 城市设施水平综合评价的统计分析[J]. 哈尔滨师范大学(自然科学版),2004,20(6):100—103
LI F Q. Statistical Analysis of Comprehensive Evaluation of Urban Facilities [J]. Journal of Harbin Normal University (Natural Science Edition), 2004, 20(6): 100—103 (in Chinese)
- [2] 张哲,汪波,杨占民. 基于主成分聚类分析的城市基础设施建设水平评价研究[J]. 西安电子科技大学学报(社会科学版),2008,18(5):27—32
ZHANG Z, WANG B, YANG Z M. Research on Evaluation of Urban Infrastructure Construction Level Based on Principal Component Clustering Analysis [J]. Journal of Xidian University (Social Science Edition), 2008, 18(5): 27—32 (in Chinese)
- [3] 谢英欣,周雨佳. K 均值法在城市基础设施建设水平分析的应用研究[J]. 统计与管理,2015,13(4):46—48
XIE Y X, ZHOU Y J. Application of K-means Method in the Analysis of Urban Infrastructure Construction Level [J]. Journal of Applied Statistics and Management, 2015, 13(4): 46—48 (in Chinese)
- [4] 张伟芳,杨永春,金建玲,等. 甘肃省城市基础设施建

- 设水平评价[J].资源与产业,2015,17(5):88—95
- ZHANG W F, YANG Y C, JIN J L, et al. Evaluation of Urban Infrastructure Construction Level in Gansu Province[J]. Resources & Industries, 2015, 17(5): 88—95(in Chinese)
- [5] 荣月静,张慧,赵显富. 基于多因素综合评价法的长三角地区城市基础设施建设水平研究[J]. 科技通报, 2016,32(2):104—110
- RONG Y J, ZHANG H, ZHAO X F. Research on the Level of Urban Infrastructure Construction in the Yangtze River Delta Based on Multifactor Comprehensive Evaluation[J]. Bulletin of Science and Technology, 2016,32(2):104—110(in Chinese)
- [6] 朱怡然. 基于基础设施建设的城市竞争力分析和比较[J]. 时代金融,2017(3):197—200
- ZHU Y R. Analysis and Comparison of Urban Competitiveness Based on Infrastructure Construction[J]. Times Finance,2017(3):197—200(in Chinese)
- [7] 何晓群. 多元统计分析[M]. 北京:中国人民大学出版社,2015
- HE X Q. Multivariate Statistical Analysis[M]. Beijing: China Renmin University Press, 2015(in Chinese)
- [8] ZOU H, HASTIE T, TIBSHIRANI R. Sparse Principal Component Analysis[J]. Journal of Computational and Graphical Statistics, 2006,15(2):265—286
- [9] ZOU H, HASTIE T. Regularization and Variable Selection via the Elastic Net[J]. Journal of the Royal Statistical Society Series B,2005,67(2):301—320
- [10] TIBSHIRANI R. Regression Shrinkage and Selection via the LASSO[J]. Journal of the Royal Statistical Society: Series B (Methodological), 1996,58(1):267—288
- [11] 喻胜华,阳林芝. 基于稀疏主成分和熵值法的中国上市银行风险评价[J]. 湖南大学学报(社会科学版), 2019,33(2):77—81
- YU S H, YANG L Z. Risk Assessment for Listed Banks in China Based on Sparse Principal Analysis and Entropy Method[J]. Journal of Hunan University (Social Science Edition), 2019,33(2):77—81(in Chinese)

Research on the Level of Urban Construction in China

CHEN Yu-xin, LIU Hui-lan

(School of Mathematics and Statistics, Guizhou University, Guiyang 550025, China)

Abstract: Based on the latest data of China Statistical Yearbook, this paper uses sparse principal component and cluster analysis methods to study the difference of urban construction level in various regions of China in 2017 and 2018. Sparse principal component analysis uses a penalty function method to sparse the loading matrix while reducing the dimensionality of data, so that the analysis results are more interpretable than principal component analysis. Moreover, cluster analysis is carried out for each region in China, and the results verify the rationality of the ranking of the sparse principal component method. The results show that the level of urban construction in China is uneven, the level of urban construction in Jiangsu, Zhejiang, Guangdong and Shandong is higher than Tibet, Ningxia and Guizhou. Finally, some suggestions are provided for the development of urbanization in all parts of China.

Key words: sparse principal component; cluster analysis; urban construction; facility level

责任编辑:李翠薇

引用本文/Cite this paper:

陈瑜鑫,刘惠兰. 我国城市建设水平研究[J]. 重庆工商大学学报(自然科学版),2021,38(4):111—117

CHEN Y X, LIU H L. Research on the Level of Urban Construction in China[J]. Journal of Chongqing Technology and Business University (Natural Science Edition),2021,38(4):111—117