Comparative Analysis of Chinese Cities' Innovation Capability Based on Spectral Clustering

Yike Li[1], Qingguo Tang[2*]

[1]School of Economic and Management, Nanjing University of Science and Technology, Nanjing, Jiangsu, 210094, China
[2]School of Economic and Management, Nanjing University of Science and Technology, Nanjing, Jiangsu, 210094, China

author Email: 1500503164@qq.com
*Corresponding author Email: tangqguo@163.com

Abstract. Starting from three aspects of city innovation environment, innovation input and innovation output, this paper constructs an evaluation index system of city innovation capability with 3 first class index and 12 second class index, and uses spectral clustering algorithm to evaluate the innovation capability of 117 first tier to third tier cities in China. The research results show that the innovation capabilities of first tier to third tier cities in China can be classified into five categories.

Keywords: City innovation capability; Evaluation index system; Spectral clustering algorithm

1 Introduction

Cities are the main components of a country, the concentration of innovation resources, and the center of national innovation. However, due to the differences in economic level, geographical location and other conditions between cities, there are also some differences in their own city innovation capability and development. Therefore, the comprehensive evaluation of city innovation capability can clearly understand the status quo of city innovation capability in various regions of China, so that relevant decision-making departments of the government can make targeted policy adjustments, which has important practical significance.

In this paper, the spectral clustering method is used to cluster the data of 12 indicators related to city innovation capability of 117 cities from the first tier to third tier cities in China in 2020, and relevant conclusions are drawn.

2 Spectral clustering algorithm

Spectral clustering is an algorithm evolved from graph theory. Its idea is derived from the spectral graph theory in graph theory. It is to transform the clustering problem into

H. Mallick et al. (Eds.): ICEMCI 2022, AEBMR 231, pp. 593–598, 2023.
https://doi.org/10.2991/978-94-6463-098-5_67

an optimal partition problem of undirected graph. The algorithm steps of standardized spectral clustering are as follows:

Input: dataset $S = (x_1, x_2, \dots, x_n)$ and number of clusters k;

Output: Cluster $A_1, A_2, \dots, A_k$.

1) Use equation (1) to calculate the similarity matrix W of $n * n$ that is, S_{ij} constitutes the similarity matrix;

$$S_{ij} = S(x_i, x_j) = \sum_{i=1,j=1}^{n} \exp \frac{-\|x_i - x_j\|^2}{2\sigma^2} \quad (1)$$

2) Use equation (2) to calculate the degree matrix D, that is, the sum of the elements in each row of W, and $n * n$ diagonal matrix composed of d_i;

$$d_i = \sum_{j=1}^{n} W_{ij} \quad (2)$$

3) Calculate Laplace matrix $L_{rsym} = D^{-1/2} L D^{-1/2} = D^{-\frac{1}{2}}(D - W)D^{-1/2}$;

4) Calculation the eigenvalues of L_{rsym} are sorted from small to large, the first k eigenvalues are taken, and the eigenvectors corresponding to the first k eigenvalues are calculated $\mu_1, \mu_2, \dots, \mu_k$;

5) Form the above k column vectors into matrix $U = \{\mu_1, \mu_2, \dots, \mu_k\}, U \in R^{n*k}$;

6) Order $y_i \in R^k$ is the vector of row i of U, where $i = 1,2, \dots, n$, and then $y_i \in R^k$ is unitized in turn, so that $|y_i| = 1$;

7) Use K-means algorithm to cluster the new sample point $Y = (y_1, y_2, \dots, y_n)$ into cluster $A_1, A_2, \dots, A_k$;

8) Output $A_1, A_2, \dots, A_k$.

3 Construction of evaluation index system

3.1 Evaluation index system

Shan constructed a system including 4 first class index and 24 second class index from the four aspects of innovation environment, innovation input, innovation output and management ability [2]. Li constructed a city innovation evaluation index system including 6 first class index and 23 second class index from three aspects: innovation environment, innovation input and innovation output [3]. Based on the research results of the two authors and the availability of data, this paper constructs a city innovation capability evaluation index system of 3 first class index and 12 second class index, as shown in Table 1:

Table 1. Evaluation index system of city innovation capability (Table credit: Original)

first class index	second class index	Unit
innovation environment	Number of ordinary colleges and universities	ten
	Number of students in ordinary colleges and universities	ten thousand
	Number of Internet broadband access users	hundred

	Number of industrial enterprises above designated size	thousand
innovation input	Number of scientific research practitioners	ten thousand
	Educational expenditure	million
	Financial expenditure on science and technology	million
	Actual utilized foreign capital	million
innovation output	Number of patent applications	item
	Number of patents granted	item
	Proportion of tertiary industry in GDP	%
	Per Capita GDP	Yuan

3.2 Source of data

The cities involved in this paper are 117 first tier to third tier cities in China. The source of each city is mainly based on the first tier to third tier cities in China described in the 2019 City Commercial Attraction Ranking released by the First Finance and Economics * New First tier Cities Research Institute. All the index data involved in this paper are from the data published on the website of the Official Bureau of Statistics by Chinese cities in 2020. Finally, because the units of each indicator data are not uniform, in order to eliminate the problem of large data differences caused by different dimensions, all indicator data are normalized.

4 Empirical analysis

4.1 Determine the number of clusters

The cluster evaluation index adopts the variance ratio criterion (Calinski-Harabasz Index) and Silhouette value. A higher Calinski-Harabasz score and Silhouette value indicate that the clustering result is better.

The spectral clustering analysis is carried out by writing codes. In order to further determine the selection of the number of clusters, the spectral clustering model is trained by substituting different values of the number of clusters. The Calinski-Harabasz score and Silhouette value are collected after each iteration calculation. The results are shown in Table 2 below:

Table 2. Calinski-Harabasz score and Silhouette value (Table credit: Original)

the number of clusters	2	3	4	5	6	7
Calinski-Harabasz score	41.349	76.346	77.375	87.952	72.721	69.123
Silhouette value	-0.036	0.156	0.291	0.812	0.170	0.168

It can be seen from the table that when clustering the city index data of China's first tier to third tier cities in 2020, with the increase of the number of clusters, the Calinski-Harabasz score and Silhouette value gradually increase and then decrease, reaching the maximum when the number of clusters is 5, which are 87.952 and 0.812 respectively.

That is to say, it is reasonable to select 5 as the number of clusters for China's first tier to third tier city index data in 2020.

4.2 Results and Discussion

The collected index data are analyzed by spectral clustering algorithm, and the results are shown in the following Table 3:

Table 3. Clustering Results of City innovation capability (Table credit: Original)

Category	City
I	Beijing, Shanghai, Shenzhen
II	Suzhou, Nanjing, Hangzhou, Qingdao, Tianjin, Guangzhou, Zhengzhou, Wuhan, Chongqing, Chengdu
III	Jinan, Hefei, Fuzhou, Nanchang, Taiyuan, Shijiazhuang Shenyang, Dalian, Changchun, Harbin, Hohhot, Changsha, Haikou, Nanning, Guiyang, Kunming, Xi'an, Lanzhou, Urumqi
IV	Wuxi, Nantong, Changzhou, Yangzhou, Taizhou, Zhenjiang, Xuzhou, Yancheng, Ningbo, Wenzhou, Shaoxing, Jinhua, Huzhou, Taizhou, Yantai, Xiamen, Quanzhou, Dongguan, Zhuhai, Foshan, Zhongshan
V	Suqian, Huai'an, Lianyungang, Zhoushan, Lishui, Zibo, Weifang, Jining, Tai'an, Weihai, Linyi, Wuhu, Bengbu, Anqing, Ma'anshan, Fuyang, Chuzhou, Zhangzhou, Putian, Ningde, Longyan, Nanping, Ganzhou, Shangrao, Jiujiang, Langfang, Baoding, Tangshan, Handan, Cangzhou, Baotou, Anshan, Qinhuangdao ,Jilin, Daqing, Huizhou, Shantou, Jiangmen, Zhanjiang, Zhaoqing, Qingyuan, Chaozhou, Meizhou, Jieyang, Luoyang, Nanyang, Xinyang, Shangqiu, Xinxiang, Yichang, Xiangyang, Jingzhou, Hengyang, Zhuzhou, Xiangtan, Chenzhou, Sanya, Guilin, Liuzhou, Mianyang, Zunyi, Xianyang, Yinchuan

It can be seen from the table that the first category is Beijing, Shanghai and Shenzhen. The common feature of these cities is that they are China's super central cities, as well as China's economic and financial centers. The second category is 10 regional central cities such as Suzhou and Nanjing. These cities have high-quality innovation capabilities and belong to high-level innovation cities. Such cities have received a lot of investment in talents and funds. At the same time, large cities have received policy support from the government, which can significantly enhance city innovation capabilities. The third category includes 19 provincial capitals such as Jinan and Hefei. These cities have strong innovation capabilities and belong to relatively high-level innovation cities. In recent years, China has strengthened the key cultivation of regional central cities, gradually forming a state of "one province, one city". This category not only improved its own innovation capability, but also driven the innovative development of surrounding cities. The fourth category includes 21 coastal cities with developed industries, such as Wuxi and Nantong, which are medium level innovation cities. The common feature of these cities is that they have relatively weak educational foundation, but their geographical location is superior, their industry is developed, their economy is developing rapidly, and they have introduced a large number of enterprises and foreign capital, so their city innovation capability has grown. The fifth category is 64 underdeveloped cities such as Suqian and Mianyang, which are low level innovation cities. In China, such cities have a large number, poor geographical location, poor economic conditions and low city innovation capability.

5 Conclusion

In this paper, the innovation capabilities of 117 first tier to third tier cities in China are classified into five categories by spectral clustering algorithm. The first category is Beijing, Shanghai, and Shenzhen, which are national central cities, and these cities have the highest innovation capability. The second category is 10 regional central cities such as Suzhou and Nanjing, which have high city innovation capabilities. The third category includes 19 provincial capital cities such as Jinan and Hefei, and the city innovation capability is above average. The fourth category includes 21 coastal cities with developed industries, such as Wuxi and Nantong, which are medium level innovation cities. The fifth category is 64 underdeveloped cities such as Suqian and Mianyang, which are low level innovation cities.

References

1. Von Luxburg U. A tutorial on spectral clustering [J]. Statistics and computing, 2007, 17(4): 395-416.
2. Shan D. Research of the construction of regional innovation capability evaluation system: Based on indicator analysis of hangzhou and Ningbo [J]. Procedia engineering, 2017, 174: 1244-1251.
3. Li X, Zheng L, Cheng J. Research on the Dynamic Evaluation System of Innovation-Driven Development of the National Independent Innovation Demonstration Zone —Based on the Quadric Weighted Dynamic Evaluation Method [J]. East China Economic Management, 2019, 33(3):7.
4. Zhou R, Liu Y, Yang Z. Spatial-Temporal Evolution and Knowledge Spillovers of Urban Innovation in China [J]. Economic Geography, 2019, 39(4):8.
5. Zhao X, Wang L. Study on the Spatial Evolution Characteristics and Influencing Factors of Innovation Agglomeration of Cities in China [J]. Economist, 2020(9):10.