The Economic Research of All Provinces in China Based on Cluster Analysis

JI Shi

School of Economics and Management, Beijing Jiaotong University, Beijing, China, 100044

Email: 18120498@bjtu.edu.cn

Keywords: cluster analysis, MATLAB, economic type

Abstract. The economic type of a province is affected by many factors, such as geographical location, GDP, population quality and so on. In this paper, MATLAB software is used to analyze the five factors of economic types, including the Gross Regional Product (billion yuan), the gross production of agriculture, forestry, animal husbandry and fishery (billion yuan), the gross output of construction industry (billion yuan), the total population (10000 people), and the total import and export volume (1000 US dollars) of the place where the business unit is located in 2018, so as to explore the development of national economy. The data comes from National Bureau of Statistics.

1. Introduction

Clustering is a process of classifying data into different classes or clusters, so the objects in the same cluster are very similar, while the objects between different clusters are very different. The goal of cluster analysis is to collect data to classify on the basis of similarity. Clustering comes from many fields, including mathematics, computer science, statistics, biology and economics. In different application fields, many clustering techniques have been developed. These techniques and methods have been used to describe data, measure the similarity between different data sources, and classify data sources into different clusters.

Cluster analysis is the general term of multivariate statistical analysis technology, which classifies research objects according to their characteristics. It directly compares the properties of different things, classifies those with similar properties into one category, and classifies those with different properties into different categories. System cluster analysis, also known as cluster analysis, is the most widely used method in cluster analysis. It quantitatively determines the similarity or kinship between samples and indicators according to the multiple indicators (variables) and multiple observation data of samples, and then links these samples or indicators to form large and small groups, forming a classification tree or icicle. Clustering analysis is simple and intuitionistic, which is mainly used in exploratory research. The results of its analysis can provide multiple possible solutions. Choosing the final solution requires the subjective judgment of researchers and subsequent analysis. No matter whether there are different categories in the actual data, the solution can be divided into several categories by using clustering analysis. The solution of clustering analysis completely depends on the choice of researchers Selecting clustering variables, adding or deleting some variables may have a substantial impact on the final solution. When using cluster analysis, researchers should pay special attention to the factors that may affect the results.

Economy is the lifeblood of a country, and maintaining economic stability is the top priority of a country. Due to the different geographical location of countries and regions, the natural resources, labor force, innovation force and other significant differences in each province, and the development of each province are also different, which will lead to imbalance. We link the cluster analysis with the unbalanced development of the country, and study whether there is unbalanced development in each province of our country through the analysis. For the provinces with unbalanced development, the national authorities should know the situation in time and take effective measures to ensure the effective operation of the macro economy of the country. Through the analysis and research of this paper, we can clearly understand the composition of China's economic types, as well as the development differences of each region, and ultimately play a

guiding role in the development of the country and the implementation of policy-making. This paper makes a cluster analysis of seven economic indicators of 31 provinces, municipalities and autonomous regions in China, i.e. Q-cluster analysis, which can objectively reflect the current economic types of various regions.

2. Problem analysis

The development of economy is very important for the development of a country. We should pay more attention to the unbalanced economic development. With the development of reform and opening up, China has entered an unprecedented period of rapid economic development, and the overall economic strength is increasing day by day. However, we should also see that the development of various regions is not balanced. Some provinces with high GDP, such as Beijing, Shanghai, Guangzhou and Shenzhen, have carried out talent introduction for many times due to their higher education level. The reason why the state attaches great importance to the higher level makes the development faster. The economic development of coastal areas is relatively rapid because of the export trade of coastal cities, flat terrain, convenient transportation and fast transportation However, due to the geographical location and transportation, the development of the central and western regions is relatively slow. Therefore, based on this situation, this paper makes a cluster analysis of the economy of all regions in the country, and makes appropriate economic policies according to the different situations of each province to ensure the balanced and effective development of China's economy.

3. Data Source

According to the data of the National Bureau of statistics, this paper selects five economic indicators of 31 provinces, municipalities and autonomous regions in 2018, as shown in Table 1.Including: the Gross Regional Product x1, the gross agricultural, forestry, animal husbandry and fishery product X2, the gross construction output value X3, the total population x4, and the total import and export volume X5 of the place where the business unit is located.

4. Data Processing

The requirement of clustering is that the samples of the same class have greater similarity, and the samples of different classes have greater difference. In order to maintain the similarity, it is often used to judge whether the distance between points is very close or relevant. As long as the distance is used to judge, it involves the operation of different indicators, and the difference in dimension and order of magnitude has a great impact on the distance judgment. In order to eliminate this impact, the data needs to be standardized before clustering.

5. Empirical Analysis

Import the data of 31 provinces, municipalities and autonomous regions in 2018 shown in Table 1 into Matlab. In order to facilitate the analysis, before clustering analysis, the standard deviation of the data is standardized, and then the standardized data is clustered. Then the similarity between variables can be calculated, the distance can be calculated, the relationship between variables can be defined, and then the clustering information can be evaluated. Finally, the clustering is created and the genealogy is made.

In the cluster analysis, the economic types of 31 provinces in China are divided into 6 categories and 1-6 categories, as shown in Table 1, the resulting genealogy is shown in Figure 1.

Table 1 cluster membership

Case	6 Clusters	Case	6 Clusters	Case	6 Clusters	Case	6 Clusters
1	1	9	1	17	2	25	2
2	2	10	4	18	2	26	2
3	2	11	5	19	6	27	2
4	2	12	2	20	2	28	2
5	2	13	2	21	2	29	2
6	2	14	2	22	2	30	2
7	2	15	3	23	2	31	2
8	2	16	2	24	2		

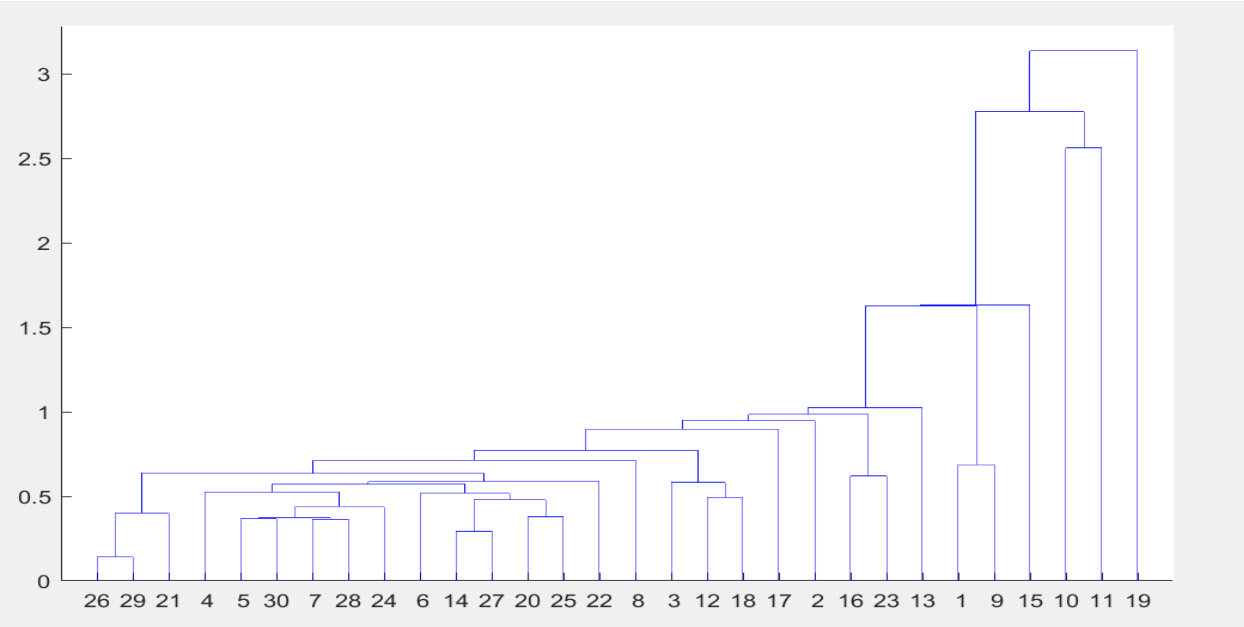

Fig.1 Cluster analysis genealogy

6. Result Analysis

Class I: Beijing, Shanghai (2)

Class II: Tianjin, Hebei, Shanxi, Inner Mongolia, Liaoning, Jilin, Heilongjiang, Anhui, Fujian, Jiangxi, Henan, Hubei, Guangxi, Hainan, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Qinghai, Ningxia, Xinjiang, Zhejiang, Hunan, Gansu (25)

Class III: Shandong Province (1)

Class IV: Jiangsu Province (1)

ClassV: Zhejiang Province (1)

ClassVI: Guangdong Province (1)

Based on the analysis of the five indicators, i.e. the Gross Regional Product (billion yuan), the gross agricultural, forestry, animal husbandry and fishery product (billion yuan), the gross construction industry output (billion yuan), the total population (10000 people), and the total import and export volume of the place where the business unit is located (1000 US dollars), 31 provinces in the country can be divided into six levels. According to the results of cluster analysis, the above classification results roughly reflect each province in the country The economic type of a province, municipality directly under the central government, or autonomous region.

Beijing and Shanghai, as the political and economic centers of the country, have outstanding comprehensive economic strength and balanced development in various fields, so they are divided into the first economic type.

Shandong Province and Jiangsu Province are relatively developed areas, which are the third and

fourth types. They have strong economic strength, convenient transportation and good development space and opportunities.

The fourth and fifth economic types are Zhejiang Province and Guangdong Province. The light industry is relatively developed and the economy develops rapidly.

Other provinces, municipalities and autonomous regions (25 in total) belong to the second type. The development of these areas is relatively slow, and even the economy of some areas is quite different from that of the first, third, fourth, fifth and sixth type provinces. This also shows that there is a great imbalance in China's economic development, and more efforts need to be invested by the state in the development of the second category of provinces.

Of course, there are also some problems in the above analysis. When selecting variables, the impact of foreign investment and tourism on the economy is not comprehensively considered. For some cities with advantages in attracting foreign investment or tourism, there is no way to comprehensively consider all the influencing factors, and the analysis results may have large deviation. But there is still a good reference value.

References

[1] Tang Pengjun, cluster analysis and discriminant analysis of the economy of all provinces in China

[2] compiled by the National Bureau of statistics of the people's Republic of China: China Statistical Yearbook 2010 [M] Beijing: China Statistical Publishing House

[3] Qi Min, Wang Fei, applied cluster analysis to the classification of economic types in some provinces and cities, School of mathematics and computer science, Yichun University, Institute of bioengineering, Yichun University

[4] Bai Zilong, Chen Qing, Chen Shaozhou. [J]. Journal of East China University of science and technology, 2001, 27 (6): 597-600

[5] Zhang Dong, Tian Yiguang, Qiao Qian et al. [J]. Journal of Wenzhou Normal University, 2001, 22 (6): 19-22