Population Classification Model of Liaoning Province Based on Cluster Analysis

Muxuan Dai

College of Mathematics, Liaoning Normal University, Dalian, Liaoning, 116029

vj162774@163.com

Abstract. Aiming at the problem of urban shrinkage in Liaoning Province, this paper established a population classification model by systematic clustering method. Based on two indicators of population contraction and GDP contraction, we defined the shrinkage rate, and classified 30 cities in Liaoning Province according to the shrinkage rate. Firstly, the weights of population contraction and GDP contraction were calculated by using the analytic hierarchy process. Secondly, the average annual growth rate of the two is weighted, and the shrinkage rate is defined as the weighted value. Then, based on the systematic clustering method, the population classification model was established by using the classical Euclidean distance, and the 30 cities were classified by SPSS software. The results of this paper on shrinking cities have important reference value for examining the future development trend of a city.

Keywords: Systematic clustering method Population data analysis SPSS.

1. Introduction

The phenomenon of urban contraction is widespread in the world, and this phenomenon has obvious regional characteristics of economic and social phenomena. This phenomenon is mainly reflected in two aspects: population reduction and economic deprivation. At present, there is no uniform definition of shrinking city, which can be measured by single population change index or multi-indicator change. According to the degree of contraction, the cities with contraction can be classified as mild, moderate and severe. It can also be divided into smart contraction and absolute contraction by referring to the changes in total population and GDP.

Since reform and opening up, China's rapid economic development, many parts of the rapid urbanization, but as the unbalanced development in different parts of the industry, many places in the city of shrinkage problem, according to the analysis of the different levels of urban central shrinkage phenomenon occurred in the northeast old industrial zone, is the traditional heavy industry city, the east developed city circle, the western frontier. As the earliest industrialized province and the national traditional heavy industry base, Liaoning is also facing this problem. Under this background, it is very necessary to investigate and study the urban shrinkage problem in Liaoning Province.

Liaoning Province has 2 sub-provincial cities, Shenyang and Dalian, 12 prefecture-level cities and 16 county-level cities, a total of 30 cities. Based on the basic data of cities in Liaoning Province, the population data of these cities and some economic development indicators of municipal and county-level cities, this paper establishes a city prediction model based on population classification to provide some solutions for the problems related to urban shrinkage.

2. Data preprocessing

Combined with the actual change rules, we preprocessed the collected population-related data as follows.

(1) The population data of Tieling and Chaoyang in 2012 were missing in Liaoning urban population statistics table. We used cubic Hermite interpolation method to solve the approximate population values of Tieling and Chaoyang in 2012, and drew the corresponding population change curve.

(2) By drawing the boxplots of the population of Liaoning cities, we found that the population data of Tieling in 2017 and 2018 were abnormal, and the cubic Hermite interpolation method was used to clean the noisy data.

(3) Some data were missing, and we used cubic Hermite interpolation method to obtain the approximate value of the missing data, which made up for the missing data.

(4) In order to cluster the two indicators of GDP and population, we calculated the annual growth rate of GDP and population of 30 cities in Liaoning Province and calculated the corresponding average value. Each indicator was given a corresponding weight, because the growth rate did not have a unit, so there was no need for standardization.

3. Population classification model

3.1. Determination of the weight of population contraction and GDP contraction

The identification indexes of shrinking cities are related to population contraction and GDP contraction. In order to compare the weights of the two factors to determine reasonable identification indexes, we established a hierarchical structure model based on analytic hierarchy process (AHP) as shown in Figure 1. The hierarchical structure model is shown in Figure 1.

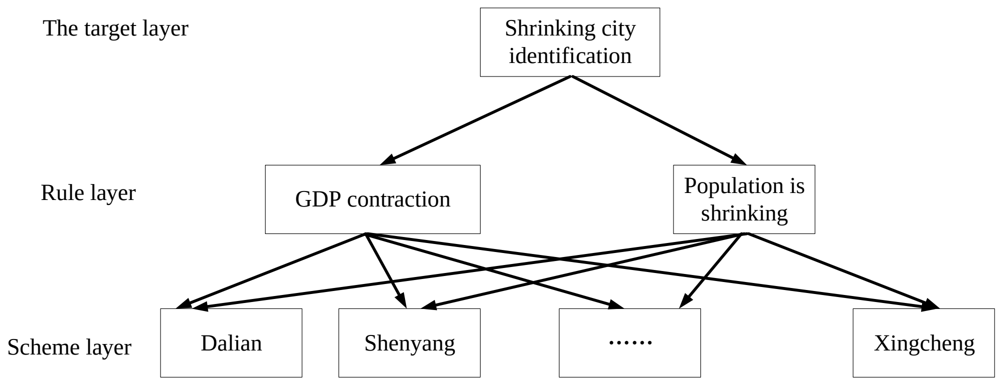

Figure 1. Hierarchical model

Combining References [1] and data analysis, population decline as a single dimension of urban shrinkage characteristics, in the multi-dimensional evaluation, should occupy more weight, at the same time for economic problems, with the industrialization and economic development of the city, there is basically no economic recession phenomenon, GDP changes should occupy less weight.

Based on the above analysis, the weight of population contraction relative to GDP contraction is defined as 7, and the following judgment matrix A is established. Since there are only two indicators, the judgment matrix A is a consistent positive reciprocal matrix [2].

$$\boldsymbol{A}=\begin{pmatrix} 1 & 7 \\ \frac{1}{7} & 1 \end{pmatrix}$$

The normalized result of eigenvector D corresponding to the maximum eigenvalue of judgment matrix A is calculated as follows:

$$\mathrm{d}=(\omega_1,\omega_2) \tag{1}$$

Get the weight of population contraction and GDP contraction, ω_1 ω_2.

3.2. Establishment of population classification model

First, we define the following indicators.

Indicators of population contraction:

$$x_i=\omega_1\overline{p_i} \tag{2}$$

GDP contraction indicator:

$$y_i = \omega_2 \overline{g_i} \tag{3}$$

Shrinkage index:

$$\beta_i = \omega_1 \overline{p_i} + \omega_2 \overline{g_i} \tag{4}$$

That is, the population contraction index is the average of the annual population rate, the GDP contraction index is the average of the annual GDP growth rate, and the contraction index is the weighted result of the population contraction index and the GDP contraction index [3].

The population classification model is established by systematic clustering method, and the algorithm is shown in Figure 2.

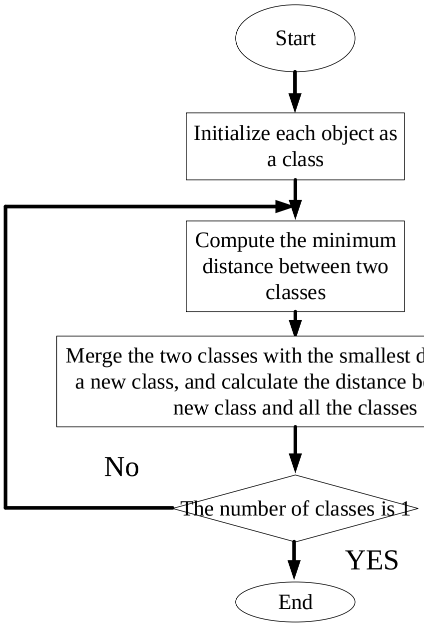

Figure 2. Flowchart of systematic clustering algorithm

Initially, all cities were divided into 30 categories, and a two-dimensional coordinate system was established, with the horizontal axis representing the population contraction index and the vertical axis representing the GDP contraction index. Classical Euclidean distance was used:

$$d_{ij} = [(x_i - x_j)^2 + (y_i - y_j)^2]^{\frac{1}{2}} = [(\omega_1 \overline{p_i} - \omega_1 \overline{p_j})^2 + (\omega_2 \overline{g_i} - \omega_2 \overline{g_j})^2]^{\frac{1}{2}} \tag{5}$$

The minimum distance between different classes was calculated.

Where, x_i, x_j represents the population contraction indicator of the two cities, and y_i, y_j represents the GDP contraction indicator of the two cities.

After the distance is calculated, the two classes with the smallest distance are merged into a new class, and the updated distances between classes are recalculated until the total number of categories is 1 [4].

3.3. Solution of the model

Step 1: According to the judgment matrix

$$A = \begin{pmatrix} 1 & 7 \\ \frac{1}{7} & 1 \end{pmatrix}$$

With the help of MATLAB's command: [v, d]=eig(A)

Where, V is the eigenvector, D is the diagonal matrix formed by the eigenvalues, and EIg is the command to find the eigenvalues and eigenvectors of the judgment matrix A.

The normalized result d=(ω_1 , ω_2) of finding the eigenvector D corresponding to the largest eigenvalue.

Finally, the weights of population contraction and GDP contraction are ω_1=0.875 and ω_2=0.125 respectively.

Step 2: Excel was used to calculate the population contraction index x_i and GDP contraction index y_i of 30 cities. The statistical data are shown in Table 1 below.

Table 1. Population indicators and GDP indicators of each city

	x_i	y_i		x_i	y_i
shenyang	0.006066	0.015677	Wafangdian city	0.009773	0.00429
dalian	0.008828	0.027517	zhuanghe	0.00906	0.0032
anshan	0.00076	0.00116	Haicheng county	0.006782	0.0064
fushun	0.004744	0.00382	donggang	0.002988	0.00327
benxi	0.003003	0.0079	fengcheng	0.006335	0.00624
jinzhou	0.006759	0.00218	linghai	0.006463	0.01186
dandong	0.005922	0.001554	Bei Town city	0.008026	0.00585
yingkou	0.008587	0.004003	Gaizhou city	0.008097	0.00661
fuxin	0.006349	0.00499	dashiqiao	0.002481	0.00518
liaoyang	0.006722	0.012454	dengta	0.009972	0.01395
Panjin	0.012188	0.05172	The soldiers mountain city	0.021473	0.00823
tieling	0.005158	0.00565	Kaiyuan city	0.004902	0.00568
Chaoyang	0.006669	0.001897	Beipiao	0.009	0.00767
huludao	0.003626	0.00282	lingyuan	0.011314	0.00245
Xinmin city	0.014264	0.00545	xingcheng	0.016064	0.00483

Step 3: Perform cluster analysis with SPSS software, and the clustering results are shown in Figure 3. [5]

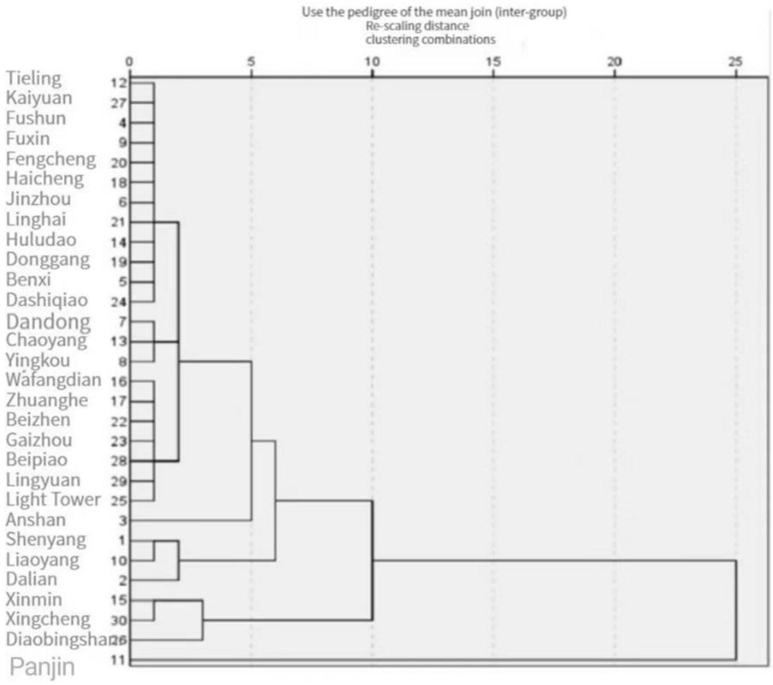

Figure 3. Pedigree of clustering results

We used Elbow's rule to determine the optimal number of clusters, defined the total distortion degree of all classes as the aggregation coefficient, and drew the following line diagram of the aggregation coefficient, as shown in Figure 4 [6].

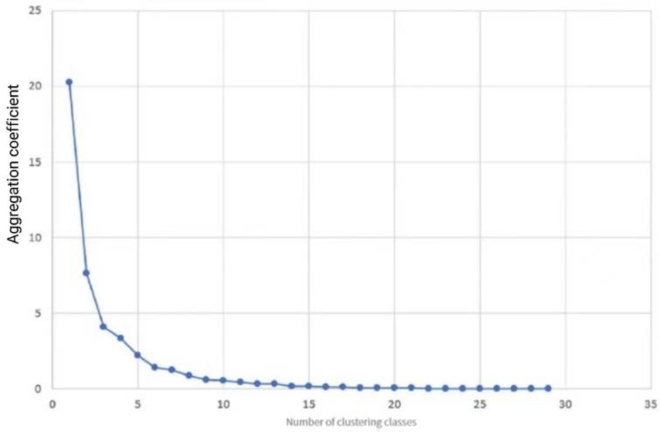

Figure 4. Curve of aggregation coefficient

It can be seen from the line diagram that the aggregation coefficient decreases significantly from class 1 to class 5. When the number of classes is greater than 5, the aggregation coefficient tends to be flat and does not change significantly. Therefore, it is feasible to divide the samples into five categories [7].

We use scatter plot to depict the general position of each category on the two-dimensional plane. The contraction index of the category near the lower left is relatively small, and the comprehensive definition of contraction index shows that the contraction degree of the category at the lower left is relatively large.Similarly, the contraction degree of the upper right group was relatively small.

In order to ensure the robustness of the results, K-means clustering method was used to perform secondary aggregation of the samples and obtain the corresponding scatter plot, as shown in Figure 6. The scatter plot drawn by the systematic clustering method was compared, as shown in Figure 5, the position difference between the two scatter plots was small, which verified the reliability of the results.

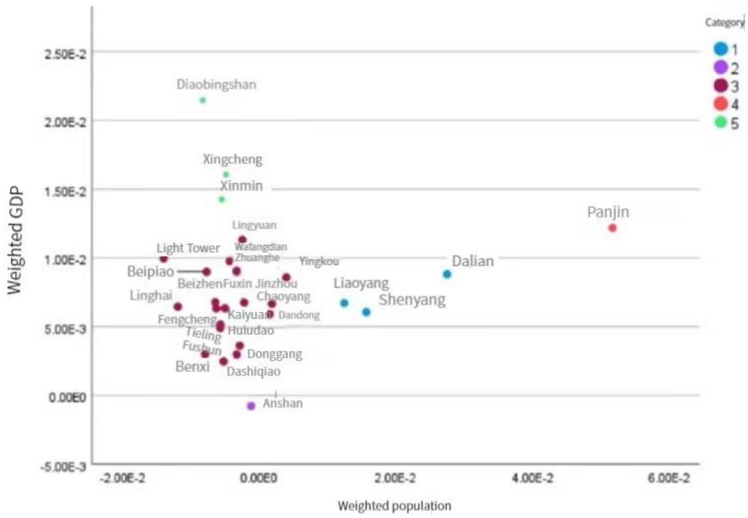

Figure 5. The corresponding scatter plot of the systematic clustering method

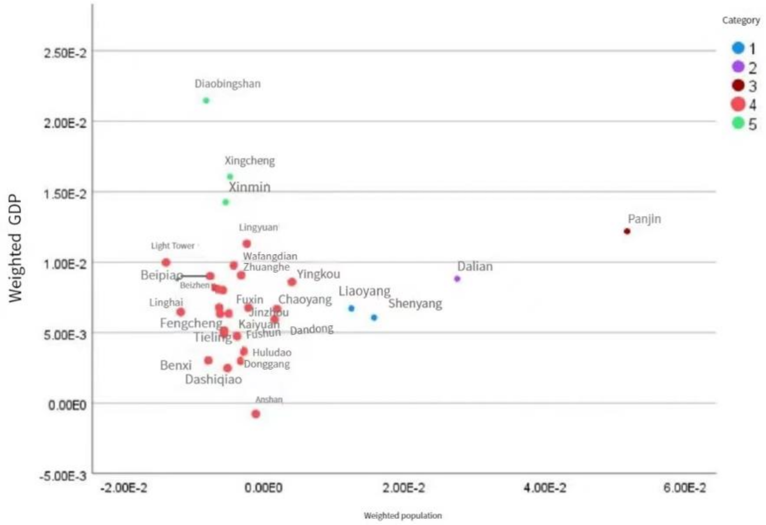

Figure 6. Scatter plot corresponding to K-means clustering method

Finally, according to the order of shrinkage rate, we defined first-level contraction, second-level contraction, third-level contraction, fourth-level contraction, and fifth-level contraction [8] (the lower the grade, the more severe the contraction) and concluded that: Anshan city for level 1 contraction, fushun, benxi, jinzhou, dandong, yingkou, fuxin, tieling, chaoyang, huludao, wafangdian, haicheng, donggang, FengCheng, LingHai, Bei Town city, gaizhou city, dashiqiao, dengta, kaiyuan city, saves, lingyuan city secondary contraction, xinmin city, adjustable soldiers mountain city, xingcheng contraction for level 3 city, Dalian, Shenyang, Liaoyang for the fourth contraction city, Panjin for the fifth contraction city. According to the known data and clustering results, the GDP and population indicators of Anshan city almost continued to decline after 13 years.As the third largest city in Liaoning Province, Anshan city has a huge population and an old heavy industry base. However, with the decline of heavy industry in northeast China and other factors, the population of Anshan city continues to lose, and it has seen continuous growth for several years. [9]

4. Conclusion

In this paper, the GDP and population indicators are integrated to complete the establishment of the shrinking city model, and provide model support for the identification of shrinking cities. This article first by writing code, analytic hierarchy process to calculate weight of the two indicators, using excel statistical and solving the two indicators of shrinkage index year by year, by SPSS software, with two indicators weight sum after average growth rate year by year as shrinkage index, USES the system clustering method in 30 cities, determine the into the category of the number, Finally, the cities are divided into five grades by calculating the contraction index of each category. In the establishment of population classification model, due to the uncertain number of classification, the system clustering method is used to overcome the defect of K-means clustering method can not determine the number of categories, and the reasonable number of categories is determined according to the aggregation coefficient. Using the classical Euclidean distance to calculate the shortest distance between categories, and giving weight to the population contraction index and GDP contraction index, the results are more reliable.

References

[1] Lu Xu, Xu Zhongzheng. [1] Recognition and formation factors of county-level city contraction in Liaoning Province. Construction of Small Towns, 2022, 40(1): 8.

[2] LI Ting. Urban clustering analysis based on real estate price and population number. Statistics and Management, 2019, (4): 71-74.

[3] [Lin Xiongbin. Measurement and influencing factors of urban contraction in China: Based on the perspective of population and economic change.] Human Geography, 2017, (1): 82-89.

[4] Qian Fengkui, Zhu Yimei, Zhang Xiaoxia, Bian Zhenxing, Shi Shuilian, Li Hongxing. [4] comprehensive measurement and influencing factors of urban shrinkage in liaoning province. China land science, 2021, 35(09): 74-83.

[5] [Teng Fei, Quan Yanying, Wang Hongyu.] The northeast three provinces population change rule and influence factors of statistical analysis [J]. Journal of tonghua normal university, 2021, and (6): 24-31. DOI:10.13877/j.carol carroll nki cn22-1284.2021.06.005.

[6] Liu Gui-Wen, XIE Fang-yun, HONG Jing-ke, Chen Chun-jiang. Based on population, economic data to the analysis of current urban contraction in China [J]. Journal of economic geography, 2019, 33(7)6: 50-57, DOI: 10.15957/j.carol carroll nki JJDL. 2019.07.007.

[7] Daniel Thomas, Casenave Fabien, Akkari Nissrine, Ketata Ali, Ryckelynck David. Physics-informed cluster analysis and a priori efficiency criterion for the construction of local reduced-order bases [J]. Journal of Computational Physics, 2022, 458.

[8] Pei Yuanzhao, Ye Linlin. Cluster analysis of MNIST data set [J]. Journal of Physics: Conference Series, 2022, 2181(1).

[9] Borlea Ioan-Daniel, Precup Radu-Emil, Borlea Alexandra-Bianca. Improvement of K-means Cluster Quality by Post Processing Resulted Clusters [J]. Procedia Computer Science, 2022, 199.

[10] de Freitas Luara Afonso, Savegnago Rodrigo Pelicioni, Menegatto Leonardo Sartori, do Bem Ricardo Dutra, Stafuzza Nedenia Bonvino, de Paz Ana Carolina Almeida Rollo, Pires Bianca Vilela, da Costa Ricardo Lopes Dias, de Paz Claudia Cristina Paro. Cluster analysis to explore additive-genetic patterns for the identification of sheep resistant, resilient and susceptible to gastrointestinal nematodes [J]. Veterinary Parasitology, 2021, 301 (prepublish).