https://doi.org/10.31449/inf.v48i12.5967 Informatica **458** (2021) 1–8 **1**

Analysis of Influencing Factors of Urban Floating Population by Clustering Algorithms

Shuhong Tang
Department of Management, Taiyuan College, Taiyuan 030032, China
Email: tangsht@outlook.com

Keywords: clustering algorithm, floating population, city, influence factor, gross domestic product, traffic jam

Received: April 1, 2024

The advancement of the economy and society has driven the increase of the urban floating population, and its analysis plays an important role in urban development. In this paper, ten influencing factors including gross domestic product (GDP) and gross industrial output value were selected to analyze the status of the floating population in ten cities such as Chengdu and Tianjin in 2010 and 2020. Two irrelevant factors were eliminated by the Pearson correlation coefficient, and the remaining eight factors were used to cluster different urban categories by an improved K-means method. The results showed that from 2010 to 2020, the increase of the floating population in Chengdu and Xi 'an was more than 100%, indicating that cities with higher GDP had a stronger ability to absorb the floating population, while high housing prices did not facilitate such absorption. These analysis results provide some references for further research on the status of the urban floating population, which can be applied in actual urban population management.

Povzetek: Razvita je bilai izboljšana K-means metoda za razvrščanje urbanih kategorij na podlagi osmih vplivnih dejavnikov, kot sta BDP in bruto industrijska proizvodnja, pri analizi urbanega migracijskega prebivalstva. Rezultati kažejo, da so mesta z višjim BDP učinkovitejša pri njihovem privabljanju, medtem ko visoke cene stanovanj negativno vplivajo.

1 Introduction

Under the impact of continuous economic development, population migration has become inevitable [1]. In the course of accelerated urbanization, the number of urban floating population is also increasing [2]. On the one hand, it provides a large labor force for urban development and stimulates urban construction and consumption [3]; on the other hand, it also brings new pressure to traffic, housing, and public security [4]. A comprehensive understanding of the status of urban floating population and a clear identification of the influencing factors of attracting the floating population can provide cities with a scientific foundation for effectively absorbing or transferring this population, which has far-reaching significance. With the progress of technology, many approaches have been applied in population research [5]. A summary of related works is presented in Table 1.

Table 1: A summary table of related works

	Approach	Result
Zhou et al. [6]	Difference and system generalized method of moments models	Young elderly people with good educational backgrounds had a positive impact on economic growth and could partially alleviate the negative impact of the overall elderly population.
Wu et al. [7]	Four latent variable analysis (LVA) approaches	Different LVA approaches all had their advantages and disadvantages when extracting behavior patterns from the aggregated population, and the employment of multiple LVA methods to find common patterns provided a more robust explanation for population dynamic changes.
Liu et al. [8]	The multi-level multinomial Logistic regression model	The employment preferences of the floating population predominantly leaned towards the traditional service industry and secondary sector, with notable spatial variations.
Chen et al. [9]	The grey model and a long short-term memory model	The elderly population would grow rapidly in the next 15 years.

From current research, it can be seen that there is relatively little study on the floating population in studies related to population changes. Literature [8] analyzed the employment characteristics of the floating population and pointed out their spatial differences. In practice, there is a significant correlation between the floating population and employment. However, there is still limited research on the factors influencing the floating population. Therefore, this paper mainly studied the influencing factors of the status of the urban floating population. Based on an improved clustering algorithm, the analysis of the status of the floating population in different cities was realized, and the relationship between different influencing factors and their absorption capacity was analyzed. This work provides some reference bases for improving the level of urban floating population absorption and the management of the floating population.

2 The status of the urban floating population and relevant influencing factors

2.1 The status of the urban floating population

The mobility of the population can be attributed to social, economic, cultural, and other motivations, such as seeking better career development and livelihood conditions, pursuing better education levels and medical services, and obtaining a better social environment and life experience [10]. According to the floating population data platform (https://www.chinaldrk.org.cn), China's floating population statistics from 1978 to 2017 are shown in Figure 1.

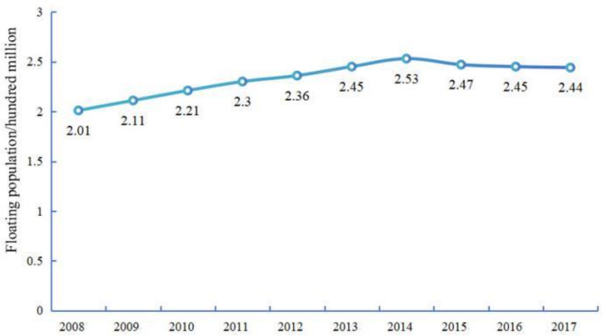

Figure 1: The status of China's floating population from 1978 to 20171

From Figure 1, it can be found that in the ten years from 1978 to 2017, the number of floating population in China has always remained above 200 million, which indicated that the number of floating population was huge. Large-scale floating population has become a basic feature of cities and a part of urban modernization that cannot be ignored. However, there are also differences in the floating population between different cities. The data from 2015 was taken as an example. The composition of the proportion of floating population is shown in Figure 2, and the proportion of floating population absorbed by different regions is shown in Figure 3.

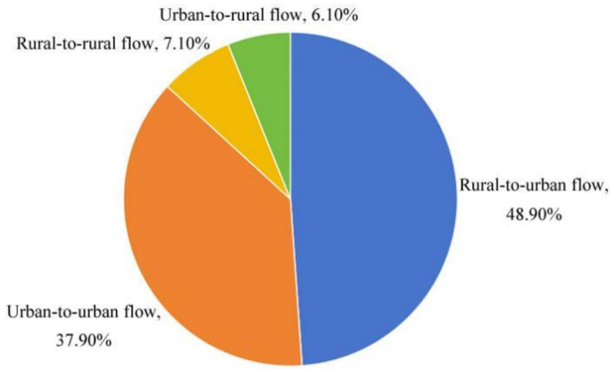

Figure 2: Proportion composition of the four types of floating population in 20152

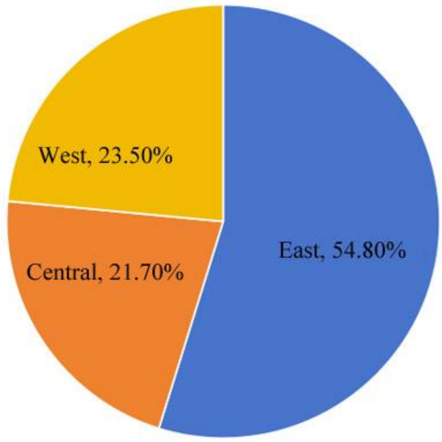

Figure 3: Proportion of floating population absorbed by different regions in 20153

Combined with Figures 2 and 3, it can be found that the floating population mainly flowed from rural to urban areas [11], exceeding a proportion of 48.9%. The proportion of the population floating from urban to urban was the second largest, reaching 37.9%. The proportion of the population floating from rural to rural and from urban to rural were small. In the statistics of different regions, the eastern region had the largest proportion of absorption, reaching 54.8%, exceeding the sum of the central and western regions. The proportion of the floating population absorbed by the west region was the smallest, only reaching 23.50%. It was concluded that the main feature of the floating population is that it flows from rural to urban areas and mainly flows to the economically developed eastern region [12].

According to the reasons for China's floating population in 2016 released by the floating population data platform, 48.55% of them are for work, followed by family migration, accounting for 22.32%, and then for business, accounting for 14.65%. From this point of view, the flow of the floating population is mainly based on economic demand. Therefore, it is bound to be related to the economic development of a city.

According to China's urban scale classification standard, supercities and megacities have a large-scale permanent population and also have a strong capacity to accommodate and absorb floating populations. Despite nearing their population limits, these cities still experience frequent population flows. Therefore, this paper mainly analyzes the situation of the floating population in supercities and megacities. This paper selects four supercities and four megacities. The data on the floating population comes from the census, as shown in Table 2.

Table 2: Status of floating population in eight cities

		2010/person	2020/person	Increase amplitude
Megacity	Chengdu	4179469	8459609	102.41%
	Tianjin	2991500	3534816	18.16%
	Wuhan	2648373	3945369	48.97%
	Shenzhen	8221754	12438738	51.29%
Supercity	Shenyang	1360148	2386818	75.48%
	Nanjing	1912592	2651812	38.65%
	Xi'an	1744754	3746945	114.75%
	Dalian	1660316	2420624	45.79%

Table 2 shows that the magnitude of floating populations is higher in megacities than in supercities. In 2010, the magnitude of floating populations in Shenzhen was more than eight million, and in 2020 it was more than 12 million, which was the largest among the eight cities studied. In terms of the increase amplitude of the floating population, Xi'an was the largest, reaching 114.75%, followed by Chengdu, reaching 102.41%, and Tianjin had the smallest increase amplitude, only 18.16%.

2.2 Analysis of influencing factors

Based on the analysis of the status of urban floating population, this paper considers the following factors:

(1) GDP, reflecting the overall economic development level of a city;

(2) gross industrial output value, reflecting the industrial development level of a city;

(3) investment in fixed assets, reflecting the investment scale of a city's fixed assets;

(4) total retail sales of consumer goods, reflecting the consumption level of a city;

(5) the average salary of employed persons, which is an important reference factor for the floating population to choose work;

(6) the number of hospitals, reflecting the medical level of a city;

(7) passenger traffic volume, reflecting the traffic level of a city;

(8) education funding, reflecting the level of investment in education;

(9) the number of students in ordinary colleges and universities, reflecting the scale of higher education in a city;

(10) the average sales price of commercial housing, which is related to the cost of renting a home for a floating population (the higher the housing price, the higher the living cost).

The relevant data in 2010 and 2020 were obtained from the statistical yearbooks and bulletin of each city. Then, they were cleaned by eliminating data with more than 70% of missing values. The other missing values were supplemented by mean values, so did the abnormal values. All the data were standardized to make the data have the same dimension:

$x' = \frac{x-min}{max-min}$,

where x' denotes the standardized data, x is the original data, max and min refer to maximum and minimum values.

The relationship between influencing factors and floating population was analyzed using the Pearson correlation coefficient [13] and Spearman correlation coefficient [14]. The correlation analysis results are presented in Tables 3 and 4.

Table 3: Pearson correlation analysis of influencing factors and floating population

Influencing factor	2010	2020
GDP	0.564**	0.567**
Total industrial output	0.212*	0.221*
Investment in fixed assets	0.133*	0.113*
Total retail sales of consumer goods	0.489**	0.492**
The average salary of employed persons	0.322*	0.484**
Number of hospitals	0.000	0.000
Passenger traffic volume	0.000	0.000
Education funding	0.141*	0.131*
The number of students in ordinary colleges and universities	0.137*	0.133*
The average sales price of commercial housing	-0.125*	-0.0127*

Note: * indicates it is significant at the confidence level of 5%; ** indicates it is significant at the confidence level of 1%.

Table 4: Spearman correlation analysis of influencing factors and floating population

Influencing factor	2010	2020
GDP	0.567**	0.577**
Total industrial output	0.375*	0.302*
Investment in fixed assets	0.215*	0.221*

Total retail sales of consumer goods	0.412**	0.433**
The average salary of employed persons	0.369**	0.418**
Number of hospitals	0.000	0.000
Passenger traffic volume	0.000	0.000
Education funding	0.142*	0.145*
The number of students in ordinary colleges and universities	0.121*	0.123*
The average sales price of commercial housing	-0.207*	-0.211*

Note: * indicates it is significant at the confidence level of 5%; ** indicates it is significant at the confidence level of 1%.

By combining the data from Tables 3 and 4, it can be observed that a stronger correlation existed between influencing factors and the floating population when the absolute value of the correlation coefficient was larger. A correlation coefficient of 0 indicated no correlation. Notably, both types of correlation analysis yielded similar results. In 2010 and 2020, there was no significant correlation found between the number of hospitals/passenger traffic volume and the floating population; therefore, these variables were excluded in subsequent studies while retaining the remaining eight influencing factors for further research.

2.3 Analysis of the clustering algorithm

To further explore the influence of various factors on the status of the floating population in the eight cities listed in Table 1, this paper classified these cities by a clustering algorithm and discussed the status of the urban floating population under different influencing factors. K-means is a method of data division [15], and its procedure is as follows.

(1) The number of target class clusters, i.e., k, was determined, and k initial clustering centers were randomly selected from n samples.

(2) The distance from each sample to k was calculated, and the samples were assigned to the nearest cluster.

(3) Cluster centers were recalculated.

(4) Steps (2) - (3) were repeated until there were no further changes in the cluster centers.

In K-means, the initial clustering center and the number of clusters have a great influence on the results [16]. To obtain better urban classification results, this paper designed an optimized K-means algorithm to enhance these two points and further improve the urban classification effect.

Firstly, the K-means++ method [17] was used in the determination of the initial clustering center. The K-means++ method improved the initial clustering centers in the K-means algorithm, which can avoid differences in clustering results due to the selection of initial points. It achieved optimization by ensuring that the distances between initial clustering centers are as far apart as possible. The details are as follows.

(1) One sample was randomly selected from N samples as the first initial clustering center c_1.

(2) The distance of each sample to c_1 was computed.

(3) The point with the largest distance was selected as c_2.

(4) Steps (2) - (3) were repeated until k initial cluster centers were selected.

For the subsequent clustering, this paper used the K-medoids method [18]. The K-medoids method addresses the sensitivity of outliers in the K-means algorithm, and it can partially alleviate the issue of local optima. Its procedure is as follows.

(1) K initial clustering centers were obtained by K-means++.

(2) The samples were divided into the nearest cluster.

(3) The cluster center was recalculated, the sum of the distance between each point and other points was calculated, and the point with the minimum value was taken as the new cluster center.

(4) Steps (2) and (3) were repeated until the cluster centers no longer changed.

In determining the number of clusters, the elbow method [19] was adopted. The principle is that the finer the sample is divided, the smaller the sum of squared error SSE. According to the curves of SSE and k, the best k value can be observed at the inflection point. The computation formula of SSE is:

$$SSE = \sum_{i=1}^{k} \sum_{p \in c_i} |p - m_i|^2,$$

where c_i is the i-th cluster, p is the sample point in the cluster, m_i is the centroid, and $|p - m_i|^2$ is the square of the distance between sample point p and center of mass m_i. The smaller the value of SSE, the better the clustering.

3 Clustering results and analysis

First, the value of k was determined according to the elbow method, and the results are presented in Figures 4 and 5.

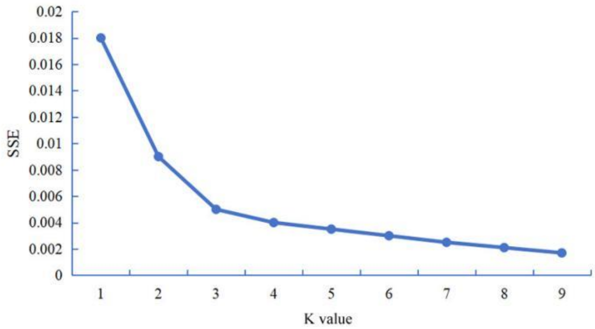

Figure 4: Determination of k value by the elbow method for data from 20104

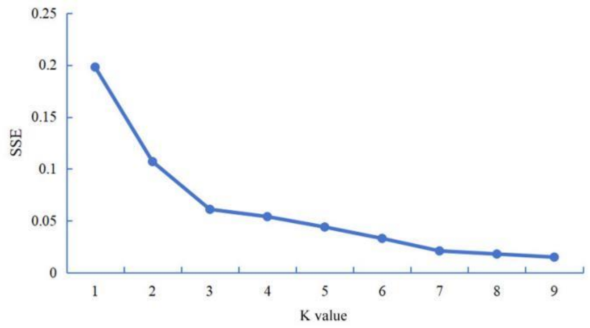

Figure 5: Determination of k value by the elbow method for data from 20205

According to Figures 4 and 5, as the k value increased, the value of SSE gradually decreased. The inflection points for the data in both 2010 and 2020 occurred at k = 3, and the change of SSE tended to be smooth as the value of k continued to increase. Therefore, it was confirmed that k = 3 could achieve a good clustering effect. In the clustering of 2010 and 2020, the value of k was determined as 3, i.e., the eight cities were divided into three classes (Table 5).

Table 5: Clustering results

2010	Category 1	Tianjin
	Category 2	Chengdu, Wuhan, Shenyang, Nanjing, Dalian
	Category 3	Shenzhen, Xi'an
2020	Category 1	Shenzhen
	Category 2	Chengdu, Tianjin, Wuhan, Nanjing, Xi'an
	Category 3	Dalian, Shenyang

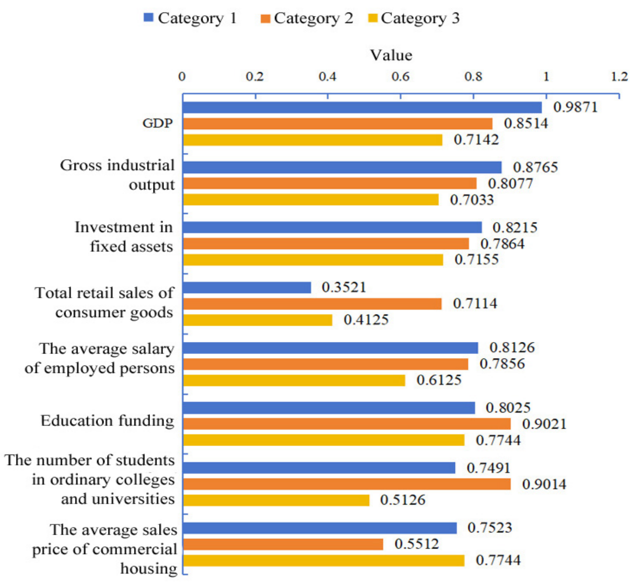

Figure 6: The clustering result analysis of year 2010 (confidence interval = 95%)

The influencing factors of different types of the urban floating population in 2010 are presented in Figure 6.

Combined with Table 5 and Figure 6, the influencing factors of different categories of urban floating population were analyzed.

(1) Category 1: Tianjin is the only city in this category. It has a high GDP, high economic development level, low consumption level, high employment income, high education level, and moderate housing price. In 2010, Tianjin's GDP level reached 683.08 billion yuan, and the total retail sales of consumer goods reached 286.02 billion yuan. The average salary of employed persons stood at 51,489 yuan, and the housing price averaged at 8,230 yuan per square meter. Under such favorable conditions, Tianjin successfully absorbed a large number of floating population.

(2) Category 2: These cities exhibit a moderate capacity to absorb a floating population, mainly characterized by a high GDP level, high consumption level, high education level, and low housing price level. In 2010, the GDP level of the five cities in Category 2 was below 600 billion yuan, which fell short of Tianjin. The primary strengths in absorbing the floating population lie in their high education level and low housing price level. Colleges and universities in Chengdu, Wuhan, Nanjing, and other cities attract a significant influx of new labor force. The average sales prices of houses in Chengdu, Wuhan, and Shenyang were all below 6,000 yuan per square meter. The average sales price of houses in Dalian was 7,044 yuan/square meter. Only Nanjing was slightly higher, reaching 9,565 yuan per square meter. The low living cost brought by the low housing price facilitates the absorption of floating population; however, due to high consumption levels as a constraint factor, these cities are comparatively less effective than Tianjin in absorbing floating populations.

(3) Category 3: These cities exhibit limited ability to absorb floating population due to their low employment income, small education scale, and high housing price level. Despite Shenzhen' high GDP level, the housing price has soared to 19,170 yuan/square meter. In addition, the number of students enrolled in ordinary colleges and universities in Shenzhen was merely 67,324 in 2010, indicating a limited appeal to the floating population. Xi'an exhibited a relatively low GDP level of only 324.15 billion yuan, suggesting an insignificant agglomeration effect, a weak capacity for generating new employment opportunities, and a poor employment attraction.

Different types of influencing factors that impact the urban floating population in 2020 are presented in Figure 7.

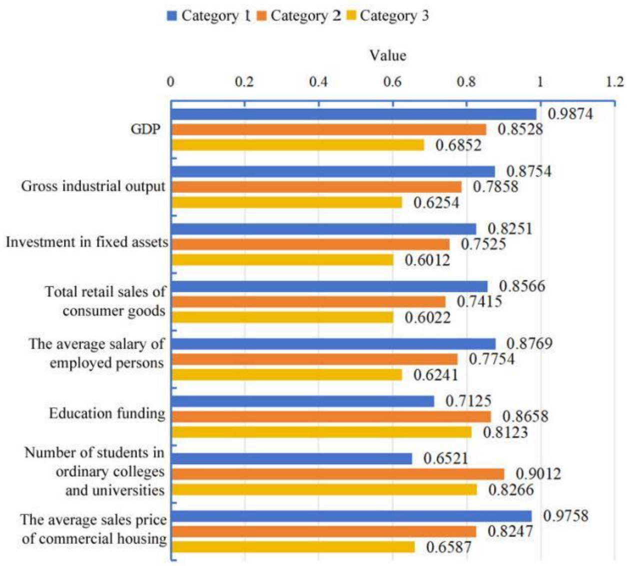

Figure 7: The clustering result analysis of year 2020 (confidence interval = 95%)1

Combined with Table 5 and Figure 7, the influencing factors of different categories of urban floating population were analyzed.

(1) Category 1: Shenzhen is the only city in this category. It has a high GDP level, high consumption level, high income level, low education scale, and high housing prices. In 2020, Shenzhen's GDP reached 2,767 billion yuan, and the average salary of employed persons reached 139,436 yuan. Furthermore, the average sales price of commercial housing reached 56,829 yuan per square meter. The number of students in ordinary colleges and universities was only 109.986 thousand, indicating a limited education scale. However, the jobs created under the high economic level and high investment intensity, and the high income level still have a strong attraction, so the floating population in Shenzhen in 2020 has exceeded 12 million.

(2) Category 2: These cities exhibit a moderate capacity to absorb the floating population. They have a high GDP level, but no more than two trillion yuan, a medium consumption level, medium employment income, a large education scale, and a medium housing price level. The GDP and consumption levels of Chengdu, Tianjin, and other cities slightly lag behind those of Shenzhen. The average salary of employed people was around 100,000 yuan. The number of students in ordinary college and universities was around 1 million, showing a large education scale. The housing price remained below 25,000 yuan per square meter. Under the high level of economic development, the large scale of education brings excellent students, and the medium consumption level and housing price level are more attractive to the floating population. Therefore, the phenomenon of population outflow is minimized while maintaining a strong capacity to absorb migrants.

(3) Category 3: This category includes Dalian and Shenyang. They exhibit a limited capacity to absorb the floating population. The floating population in 2020 was less than 2.5 million. They are characterized by low GDP, consumption level, income level, and housing prices, suggesting that these cities have a low level of economic development, insufficient agglomeration effects, limited investment and employment opportunities, and lack of competitiveness. Despite possessing a moderate level of education, college graduates exhibit limited inclination to remain and work locally. The average salary of employed people was below 100,000 yuan, which is insufficient to attract the floating population.

4 Discussion

The floating population is an important part of the population, and the increase in the floating population generally contributes to improving the economic level of cities. However, a large influx of floating population may also have certain impacts on city infrastructure, medical standards, etc. Therefore, managing the floating population is an important aspect of urban population management. Due to their large numbers and rapid changes, there is currently limited research on the floating population and analysis of different factors affecting their conditions. In order to provide theoretical support for urban population management, this study analyzed the influencing factors on the status of the floating population based on clustering algorithms using data from 2010 and 2020.

From the clustering results, it can be observed that the improved k-means method effectively compensated for the shortcomings of the traditional k-means algorithm and demonstrated better performance in clustering. Based on the elbow method, the number of clusters was determined to be 3. Combining the analysis from Figures 6 and 7, it can be concluded that cities with higher GDP levels and employee wages had a greater attraction to floating populations, while cities with higher housing prices and consumption levels had a smaller attraction to floating populations. This result aligns with reality.

In summary, Xi'an's GDP has grown from 324.1 billion yuan in 2010 to 1,002 billion yuan in 2020, more than tripling in size. As the most economically advantageous city not only in Shaanxi but also in Northwest China, Xi'an accounted for 38% of Shaanxi's total economic output in 2020. The city boasts well-developed urban functions, infrastructure, and education and healthcare systems. Xi'an Xi'an has emerged as the preferred destination for a substantial number of migrant workers. Therefore, in the ten years, Chengdu has significantly enhanced its capacity to attract the floating population, leading to a remarkable increase of 114.75% in their numbers. During this period, Chengdu's population attractiveness and agglomeration level have been consistently rising. As the provincial capital city, Chengdu has a strong ability to attract the population of surrounding cities, and its rich university resources have attracted a large number of talent inflows. In contrast, Tianjin entered the middle and late stage of urbanization in 2020, and its economic development is relatively sluggish. Instead of increasing, its GDP is decreasing, and its absorption capacity for the floating population is also declining. In general, the city's economy, employment, education, housing prices, and other factors will have an influence on the situation of the floating population.

The present study conducted a certain analysis on the influencing factors of urban floating population and obtained some reliable results. However, there are still some shortcomings, such as only studying the data from 2010 and 2020, insufficient selection of influencing factors, and lack of analysis on the floating population situation in all cities in China. Therefore, in future work, it is necessary to consider more influencing factors and conduct cluster analysis on the floating population situation in all cities to better optimize the current level of management for urban floating population.

5 Conclusion

This paper mainly analyzed the factors that impact the urban floating population and designed an improved K-means method for city clustering. It was found that cities with higher levels of economic development, higher employment income, and lower housing price levels tend to attract more floating population, but different combinations of these influencing factors can also have varying effects on the capacity to absorb the floating population. The results of the analysis can provide some theoretical references for urban population management. According to the changes in local city's GDP level and housing price level, different measures can be taken to absorb or restrict the floating population. For example, for western cities, attracting the floating population can be achieved by increasing salary levels and expanding labor demand. In addition, timely adjustments of relevant policies based on economic growth and demand changes can further promote reasonable population mobility.

References

[1] Rushiti M, Skenderi F, Hamiti R (2019). Characteristics of the natural movement of the population in Republic of North Macedonia, 2018. *Knowledge International Journal*, 34(6), PP. 1755-1760. https://doi.org/10.35120/KIJ34061755R.

[2] Xu Y, Wu D (2019). Public Service Delivery of Floating Population from the Perspective of Governance: A Case of Hangzhou. *2019 16th International Conference on Service Systems and Service Management (ICSSSM), Shenzhen, China*, pp. 1-6, https://doi.org/10.1109/ICSSSM.2019.8887605.

[3] Mosikari TJ, Eita JH (2020). CO2 emissions, urban population, energy consumption and economic growth in selected African countries: A Panel Smooth Transition Regression (PSTR). *OPEC Energy Review*, 44(3), pp. 319-333. https://doi.org/10.1111/opec.12184.

[4] Roy S, Cooper D, Mucci A, Sana B, Chen M, Castiglione J, Erhardt GD (2020). Why is traffic congestion getting worse? A decomposition of the contributors to growing congestion in San Francisco-Determining the Role of TNCs. *Case Studies on Transport Policy*, 8(4), pp. 1371-1382. https://doi.org/10.1016/j.cstp.2020.09.008.

[5] Wang Y, Wang Z, Zhou C, Liu Y, Liu S (2020). On the Settlement of the Floating Population in the Pearl River Delta: Understanding the Factors of Permanent Settlement Intention versus Housing Purchase Actions. *Sustainability*, 12(22), pp. 1-20. https://doi.org/10.3390/su12229771.

[6] Zhou X, Ting LI (2019). The influence of the elderly population on economic growth in China. *The Singapore Economic Review*, 66(06), pp. 1595-1611. https://doi.org/10.1142/S0217590818420080.

[7] Wu L, Chikaraishi M, Nguyen HTA, Fujiwara A (2021). Analysis of post-disaster population movement by using mobile spatial statistics. *International Journal of Disaster Risk Reduction*, 54(29), pp. 102047. https://doi.org/10.1016/j.ijdrr.2021.102047.

[8] Liu Z, Qi W, Liu SH, Qi HG, Jin HR, Zhang XF. Employment choice of the floating population and influencing factors in China. *Progress in Geography*, 42(6), pp. 1055-1068. https://doi.org/10.18306/dlkxjz.2023.06.003.

[9] Chen HQ, Guo GC, Qin CX, Li ZB (2021). Research on Elderly Population Prediction Based on GM-LSTM Model in Nanjing City. *Computer Science*, 48(6A), pp. 231-234. https://doi.org/10.11896/jsjkx.200900142.

[10] Meng YY, Zhang XD, Wang J (2021). Lock-in" and "Pull-back": The Impact of Medical Insurance on the Residence Intention of Floating Population. *Journal of Northeastern University (Social Science)*, 23(4), pp. 67-75. https://doi.org/10.15936/j.cnki.1008-3758.2021.04.009.

[11] Ke WQ, Xiao BY, Lin LY, Zhu Y, Wang Y (2023). Interprovincial urban and rural floating population evolution of China and its relationship with regional economic development. *Acta Geographica Sinica*, 78(8), pp. 2041-2057. https://doi.org/10.11821/dlxb202308012.

[12] Shi Q, Liu T (2019). Glimpsing China's future urbanization from the geography of a floating population. *Environment and Planning*, 51(4), pp. 817-819. https://doi.org/10.1177/0308518X19834572.

[13] Jose-Alfonso AC, Anibal ZC, Carolina TB, Roberto EO, Carlos-Eduardo ZC (2019). A Pearson Correlation Analysis of the SoftwareEngineering Practice in Micro and Small-SizedSoftware Industry of Sinaloa, Mexico. *IEEE Latin America Transactions*, 17(2), pp. 210-218. https://doi.org/10.1109/TLA.2019.8863166.

[14] Xiao C, Ye J, Esteves RM, Rong C (2016). Using Spearman's correlation coefficients for exploratory data analysis on big dataset. *Concurrency and Computation: Practice and Experience*, 28(14), pp. 3866-3878. https://doi.org/10.1002/cpe.3745.

[15] Shrestha R, Kongprawechnon W, Leelasawassuk T, Wongcumchang N, Kondo T (2020). Intensity and K-Means Clustering based Treatment Classification of Posterior Capsular Opacification. *2020 6th*

International Conference on Control, Automation and Robotics (ICCAR), Singapore, pp. 541-545. https://doi.org/10.1109/ICCAR49639.2020.9107985.

[16] Rosyid H, Mailok R, Lakulu MM (2019). Optimizing K-Means Initial Number of Cluster Based Heuristic Approach: Literature Review Analysis Perspective. *International Journal of Artificial Intelligence*, 6(2), pp. 120-124. https://doi.org/10.36079/LAMINTANG.IJAI-0602.40.

[17] Aggarwal S, Singh P (2023). Software fault prediction using hybrid swarm intelligent cuckoo and bat-based k-means++ clustering technique. *International Journal of Advanced Intelligence Paradigms*, 25, pp. 341-359. https://doi.org/10.1504/ijaip.2021.10016288.

[18] Ramdhan W, Sitompul OS, Nababan E, Nasution S (2022). Clustering Algorithm-Based Pandemic Multicluster Framework Analysis: K-Means and K-Medoids. *2022 IEEE International Conference of Computer Science and Information Technology (ICOSNIKOM)*, pp. 1-6. https://doi.org/10.1109/ICOSNIKOM56551.2022.10034927.

[19] Jaafar BA, Gaata MT, Jasim MN (2020). Home appliances recommendation system based on weather information using combined modified k-means and elbow algorithms. *Indonesian Journal of Electrical Engineering and Computer Science*, 19(3), pp. 1635-1642. https://doi.org/10.11591/ijeecs.v19.i3.pp1635-1642.