Evaluation of Teacher Allocation in Regional Compulsory Education Industry Based on K-means clustering and GIS Spatial analysis: A Case Study of teacher-student ratio

Limin Liu[1,2], Yu Chen[1*], Li Yang[1], Ke Yan[1]

[1]College of Computer, Hubei University of Education, Wuhan, 430205, China
[2]Hubei Education Cloud Service Engineering Technology Research Centre, Hubei University of Education, Wuhan 430205, China
*Corresponding Author.

Abstract

The allocation of teacher resources can reflect the situation of regional education disequilibrium. It is an important research content to effectively evaluate the differences of regional teacher resource allocation and reveal the characteristics of spatial location. Take teacher-student ratio of compulsory education of five adjacent cities in China as an example, this paper focus on difference analysis and balanced evaluation of teacher resource allocation. Combined with geographical location, Differentiation coefficient, K-means clustering and GIS hot spot analysis methods are used to carry out the classification of teacher resource allocation and the spatial aggregation model of cold and hot spots. The results show that: there is an imbalance between different cities, as well as between districts and counties in same city.The results of K-means clustering and GIS hot spot analysis also show that regional teacher resource allocation has the characteristics of numerical classification and spatial aggregation. The k-means algorithm can aggregate the teacher-student ratio of 30 districts into three categories, revealing the category characteristics of teacher allocation. The results of GIS hotspots show that there are high value aggregation and low value aggregation in teacher configuration. The methods and research results of this paperprovide newresearch ideas for resource balance evaluation and spatial pattern analy sis of education industry.

Keywords: Education industry, Teacher-student ratio,data mining, K-means, GIS hotspot analysis

I. Introduction

With the rapid development of Chinese compulsory education, effectively evaluating the current situation and characteristics of teacher allocation in the compulsory education stage is of great significance for the further realization of education equity.

In recent years, many scholars have done a lot of research on the allocation of basic education teachers [1-4] and the differences between urban and rural areas [5]. Through observation and investigation, Mauro Tamayo et al. studied the urban-rural distribution differences of Inclusive Education in Chile [1], and found that the urban-rural differences were obvious. Zhang et al. analyzed the investment in education funds, teaching staff and school running conditions, and found that the regional and urban-rural education gap in the stage of compulsory education was significantly narrowed [6]. ETomulcalculated Gini index to determine the relationship between inequality in education in Turkey[7] and the average years of schooling. Peng et al. analyzed the relationship among child population, economic development and regional inequality [8], and the result shows there is great inequality of educational resources between different provinces in China due to unbalanced economic development. Li used Gini coefficient and range rate to empirically test the balanced development of China's compulsory education from the perspective of per student funding [9], and found that the Gini coefficient of per student funding showed a downward trend, with obvious regional differences. Shen analyzed the student teacher ratio, gender ratio, educational background distribution and caste distribution of basic education in India [10], and found the characteristics of resource allocation differentiation. Zhou studied the imbalance of teachers' age structure,

educational background structure and professional title structure in primary and secondary schools in Wuhan [11], and put forward some optimization suggestions.

In addition, in view of spatial characteristics of educational resources, researchers conduct a series of studies based on GIS technology [12-16]. Zhang [12] constructsequilibrium analysis indicators for basic education based on educational opportunities, resource allocation, and educational quality and derives school distance and accessibility through GIS overlay analysis and network analysis. RG Taylor et al.[13]adopt GIS spatial analysis technology to study school distribution and layout adjustment plan in Johnston County, North Carolina. Wu et al. [14]study distribution and spatial pattern of educational facilities in Hangzhou, and evaluate fairness of educational resources. Guo et al. investigates distribution characteristics and layout efficiency of educational resources in towns/streets, finding significant spatial distribution clustering characteristics [15]. Although the above researcheshave done a lot in accessibility analysis and evaluation of educational resource layout, there is insufficient research on the law of educational development, the pattern of resource spatial layout, and the evolution law [16].

In view of this, this papertakes teacher-student ratio indicator to study the 2017 compulsory education stage in 5 spatially adjacent cities in China based ondifference coefficient, K-means clustering [17] and GIS hotspot analysis method.Through data mining and spatial analysis, allocation of teacher-student ratios in different regions is divided by clustering to analyze cold and hotspot spatial aggregation modes of allocation of analyst-student ratios, thereby providing new research ideas for the evaluation of regional teacher resource equilibrium and spatial pattern analysis.

II. Study Site and Datasets

This paper selects five spatially adjacent cities in central China, and the data of teacher-student ratios in different stages of compulsory education in 2017 for evaluation and equilibrium research.

The five cities are named as City1, City2, City3, City4, City5, and City1 is the provincial capital city. Where, City1 has jurisdiction over 13 districts and counties; City2 has jurisdiction over 7 districts and counties; City3 has jurisdiction over 3 districts and counties; City4 has jurisdiction over 6 districts and counties; and City5 has jurisdiction over 1 district.

Naming of districts and counties is internally sorted based on the city and provinceof its location. For example, the districts and counties under the jurisdiction of City1 are named as: City1-1, City1-2, etc. For example, the districts and counties under the jurisdiction of City2 are named as: City2-1, City2-2, etc. The research area is divided into two research scales: city and district/county.

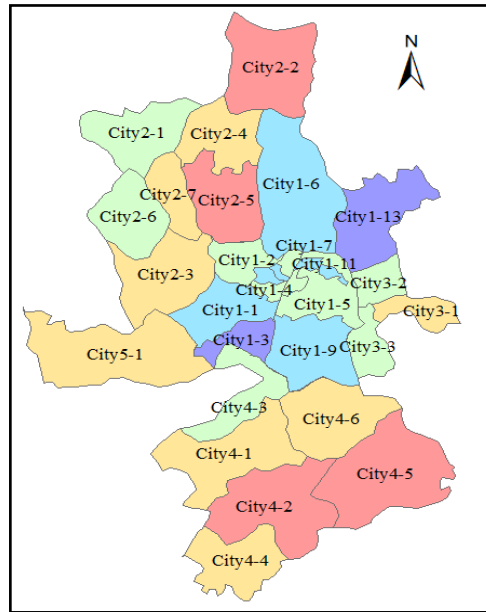The spatial distribution of the five cities in the study area is shown in Fig.1.

Fig 1: Spatial distribution of 30 districts and counties of the five cities

According to the standard, the teacher-student ratio of primary school, junior high school and senior high school is no less than: 1:23.8,1:19.3 and1:15.8.

Table 1 Teacher student ratio of compulsory education in five cities in 2017

Cities	Teacher student ratio in primary schools	Teacher student ratio in junior high school	Teacher student ratio in high school
City1	1: 18.4	1: 10.7	1: 9.2
City2	1: 15.3	1: 10.0	1: 13.1
City3	1: 14.0	1: 11.0	1: 13.3
City4	1: 19.8	1: 14.7	1: 13.9
City5	1: 17.1	1: 9.6	1: 12.6

Table 2 Teacher student ratio of compulsory education in 30 districts and counties

Districts and Counties	Teacher student ratio in primary schools	Teacher student ratioin junior high school	Teacher student ratio in high school
City1-1	1: 18.4	1: 9.0	1: 8.0
City1-2	1: 19.5	1: 12.0	1: 10.4
City1-3	1: 14.9	1: 7.9	1: 6.3
City1-4	1: 19.4	1: 12.4	1: 10.7
City1-5	1: 20.4	1: 11.2	1: 11.2
City1-6	1: 19.9	1: 10.3	1: 8.2
City1-7	1: 17.5	1: 11.0	1: 10.2
City1-8	1: 18.2	1: 10.8	1: 9.3
City1-9	1: 17.1	1: 10.5	1: 9
City1-10	1: 18.3	1: 11.5	1: 8.9
City1-11	1: 21.6	1: 12.9	1: 9.1
City1-12	1: 16.2	1: 11.5	1: 10.6
City1-13	1: 16.7	1: 8.4	1: 6.4
City2-1	1: 14.6	1: 8.9	1: 10.8
City2-2	1: 15.5	1: 9.5	1: 14.4
City2-3	1: 16.6	1: 9.6	1: 12.3

City2-4	1： 16.0	1： 10.6	1： 12.5
City2-5	1： 16.2	1： 13.0	1： 16.1
City2-6	1： 11.8	1： 7.8	1： 11.4
City2-7	1： 14.5	1： 10.4	1： 12.8
City3-1	1： 14.5	1： 11.8	1： 13.9
City3-2	1： 13.4	1： 9.0	1： 11.4
City3-3	1： 13.1	1： 10.5	1： 11.6
City4-1	1： 21.3	1： 11.5	1： 13.7
City4-2	1： 18.8	1： 17.8	1： 16.9
City4-3	1： 15.0	1： 8.9	1： 11.0
City4-4	1： 19.7	1： 15.7	1： 13.6
City4-5	1： 19.2	1： 16.4	1： 14.7
City4-6	1： 22.6	1： 17.1	1： 13.7
City5-1	1： 17.1	1： 9.6	1： 12.6

III. Methods and Analysis

3.1 K-means clusteringmethod

K-means (k-means clustering algorithm) algorithm is a widely used unsupervised clustering algorithm [18,19], which is simple to implement and has a good and stable clustering effect. To master the allocation characteristics of teacher-student ratio in 30 districts and counties, this paper takes the deviation allocation standard of the teacher-student ratio of the 3 learning stages of each district/county as the deviation degree to construct a 3-dimensional feature attribute, and uses K-means algorithm to cluster samples from 30 districts and counties and find the classification characteristics.

Based on k random pointsformed by the initial cluster center,K-means algorithm is to calculate the distance from each point to the cluster center, and merge the sample to the closest cluster midpoint in view of the distance. New cluster centers are calculated based on the new class following one-time iteration. After multiple iterations, if the difference between the two adjacent cluster centers changes very little, it means that the sample clustering ends and the clustering criterion function converges. When the amount of data increases, small memory consumption is required, and when the density of spatial clustering is not uniform, the clustering quality is fine.

For a m dimensional observation sample $\{x_1, x_2, ... x_n\}$ in space, set the sample classification category to k. Randomly select k objects as the center point of the initial cluster, and use Euclidean distance as the similarity measure. By finding the minimum variance, the similarity between each sample and the cluster center $\{\mu_1, \mu_2, ... \mu_k\}$ is calculated. The variable $\gamma_{ij} \in \{0,1\}$ indicates whether the i-th sample is clustered into the k-th class. If it is true, then $\gamma_{ij}=1$. Otherwise, $\gamma_{ij}=0$. Update the cluster centers and sample categories through multiple iterations until the function converges.

In terms of clustering effect of this algorithm, each sample can only belong to one class so that the similarity within the class is as big as possible, and the difference between the classes is as big as possible. The optimization goal of K-means is:

$$J(\gamma, \mu) = \sum_{i=1}^{n} \sum_{j=1}^{k} \gamma_{ij} \left\| x_i - \mu_j \right\|_2^2 \qquad (1)$$

3.2 GIS Hotspot analysis method

Combining geospatial information, spatial autocorrelation and GIS hotspot analysis methods are taken to identify the spatial location characteristics of teacher allocation at the district and county scale, thus understandingwhether teacher allocation is sufficient in spatial clustering.

GIS hotspot analysis [20] is a spatial clustering mode analysis method. Based on a given set of weighted elements, this method is mainly to use local General G index statistics to identify statistically significant hot and cold spots. Hotspot analysis will look at every element in the environment of adjacent elements. Therefore, only an isolated high value will not constitute a hotspot. Single elements and its surroundings all belong to high values.That is, the area is an aggregation area of high values, thus acquiring the name hotspot. Conversely, a cold spot means not only its own value is very low, but also its adjacent values are low, namely a cluster area of low values. In the analysis of GIS hotspots based on the teacher-student ratio in this paper, hotspots represent areas where teachers are abundant, and cold spots represent areas where teachers are relatively scarce.

The model calculation returns two indicators of Z score (Gi_Zscore) and P value (Gi_Pvalue). Z score represents the multiple of the standard deviation, while P value represents the probability that the observed spatial pattern is created by a random process. Through the combination of Z score and P value, statistically significant clusters or discrete patterns are found.

IV. Results and Discussion

4.1 The overall situation of in the study area

4.1.1 The standard reaching of teacher student ratio
According to the allocation standards of teacher-student ratio in different compulsory education stages, the allocation compliancestatus in 2017of the 5 cities and provinces and 30 districts and counties under the jurisdiction of the study area are analyzed.

According to the data in Table 1, the overall allocation is relatively fine at the city, province, and district/county scales. On the city-province scale, the allocation of teacher-student ratio of the five cities are all up to the standard in the primary, junior middle, and high school stages, but there are certain differences in the degree of allocation abundance between different cities and provinces. For example, the student-teacher ratio of City3 is 1:14 in primary school, while that of City4 is 1:19.8 in primary school. As a provincial capital, City1 only has the best teacher-student ratio allocationin high school among the five cities and provinces. At the district/county scale, the data from the 30 districts and counties in Table 2 show that only City4-2 under the jurisdiction of City4does not meet the standard in high school stage, and the remaining learning stages are all up to the standard; different districts and counties have obvious differences in allocations, showing differences between the districts and counties under the jurisdiction of different cities and provinces.

4.1.2 Difference coefficients of teacher student ratio
To evaluate allocation equilibrium of cities, provinces and districts, counties under its jurisdiction, this paper makes calculations based ondifference coefficient.Under greater difference coefficient, data dispersion is higher and allocation disequilibrium is more obvious. In this paper, difference coefficient is calculated based on the model proposed by Pearson (K.) by dividing the standard deviationby the average.

The difference coefficients forprimary, junior middle and high school stages in the five cities and provinces are: 0.14013, 0.152059, and 0.184333, respectively. The results suggest that high school has the most obvious allocation disequilibrium among the five adjacent cities and provinces, showing the most significant difference.

The difference coefficients for the districts and counties within the five cities and provinces are shown in Table 3. The results indicate that except City5, the 5 cities show significant equilibrium in internal allocation; the 6 districts

and counties under the jurisdiction of City4 have the most significant disequilibrium in teacher-student ratio allocation in junior middle school, which should arouse concern; the 13 districts and counties under the jurisdiction of the provincial capital City1 also show obvious disequilibrium in teacher-student ratio allocation in high schools, demonstrating significant differences in allocation of teachers.

Table 3 Analysis of imbalance among districts and counties in each city

districts and counties	Difference coefficients in primary school	Difference coefficients in junior high school	Difference coefficients in high school
thirteen districts and counties in City1	0.103648	0.156082	0.197305
seven districts and counties in City2	0.124756	0.156565	0.130997
three districts and counties in City3	0.050519	0.13714	0.103545
six districts and counties in City4	0.149832	0.303931	0.140395
one district in City5	0	0	0

4.2 K-means clustering of teacher-student ratio allocation in 30 districts

According to the description in Section 3.1, at the district/county scale, a feature vector is constructed based on theteacher-student ratio allocation results of primary, junior middle, and high school stages of 30 districts and counties, and the K-means algorithm is used for clustering to find the category classification features of allocation in district/county.

By selecting the number of initial cluster centers K, it is possible to test multiple different cluster centers. According to the sum of squared errors of different K values, determine the value with the fastest slope change, that is, the optimal K value. The optimal K value in this paper is 3, that is, the teacher-student ratio data of 30 districts and counties will be divided into 3 categories according to the differences in distribution. The clustering results are shown in Figure 2. The 3-dimensional space is composed of three attributes used for clustering: primary school teacher-student ratio, junior middle school teacher-student ratio, and high school teacher-student ratio. Each scattered point represents a district, with different colors and shape logos representing different categories. From the clustering results, it is found that City1-1, City1-3, City1-6 and City1-13 have been divided into category 1; City1-2, City1-4, City1-5, City1-7, City1-8, City1-10, City1-12, City2-4, City2-5, City3-1, City4-1, City4-2, City4-4, City4-5, City4-6 are divided into category 2; other districts and counties are clustered into categories 3. The districts and counties within the category have similar allocation of teachers in the three learning stages, while great differences exist in allocation of teachers in different districts and counties. From the clustering results, it is found that the districts and counties under the jurisdiction of City1 are mainly clustered into categories 1 and 2, indicating unbalanced allocation within the city. Moreover, some districts and counties have allocation characteristics similar to City4. The districts and counties under the jurisdiction of City2 and City3 are mainly classified into category 3, indicating relatively similar allocation characteristics of the two cities in overall.
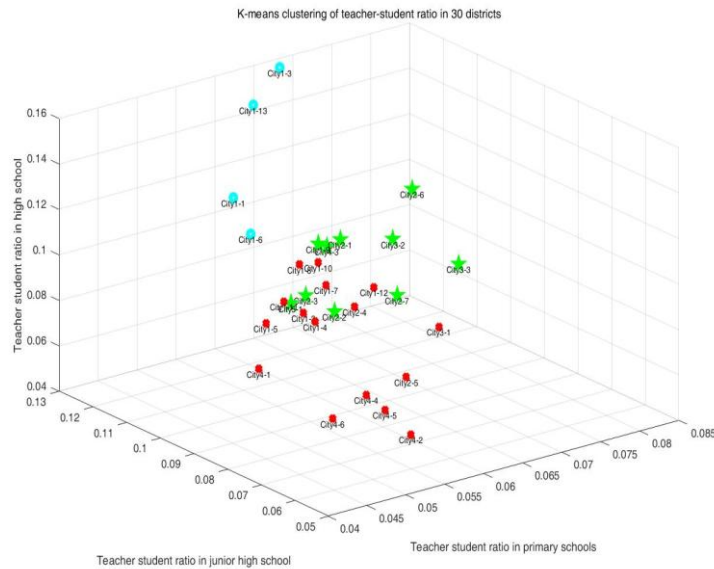
Fig 2: K-means clustering of teacher-student ratio in 30 districts and counties
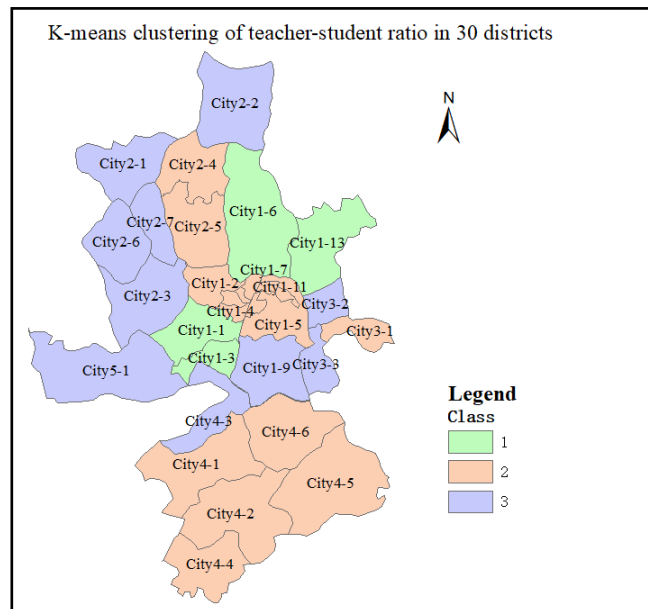

Fig 3: Spatial distribution of K-means clustering results

4.3 GIS hotspot analysis model clustering

Based on the description in 3.2, to obtain teacher-student ratio allocation in the 30 districts and counties, we spatially aggregate the characteristics and location characteristics; analyze whether the provincial capital has a siphon effect on surrounding cities, and use GIS hotspot analysis method for spatial mode clustering of primary school, junior middle school and high school stages of compulsory education, thus revealing the allocation and spatial distribution patterns of teachers, including the spatial distribution patterns of high-value clustering points and land-value clustering points, etc.

GIS hotspot analysis is made on the teacher-student ratio of primary school, junior middle school and high school respectively, with results shown in Figure 4-6. Figure 4 indicates that the three districts of City2-1, City2-6, and City2-7 are spatially adjacent to each other, which are high-value spatial aggregation areas in teacher-student ratio

allocation of primary school stage, indicating sufficient teachers in this area in overall; City1-2, City1-7 , City1-10, City1-4 are spatially adjacent to each other, displaying characteristics of low-value spatialaggregationin teacher-student ratio; City4-2 and City4-5 also exhibit low-value aggregation in space; other areas show insignificant spatial correlation.
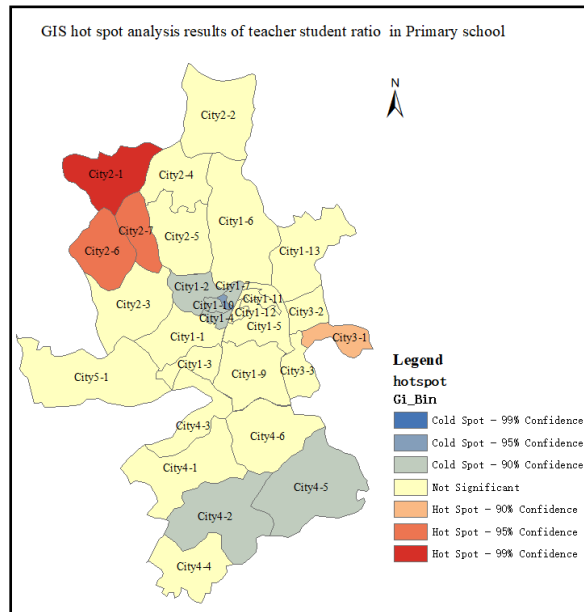

Fig 4: Hotspots analysis results of primary teacher-student ratio in 30 districts

Hotspot analysis results of teacher-student ratio in junior middle schools in Figure 5 indicate that: City2-1 and City2-3 exhibit high teacher-student ratio in spatial clusters, and this area has sufficient teachers as a whole; City4-2, City4-4 and City4-5 exhibit low-value spatial aggregation characteristics, indicatinglow teacher-student ratio andinsufficient allocation in this area,while other districts and counties show insignificant spatial correlation.
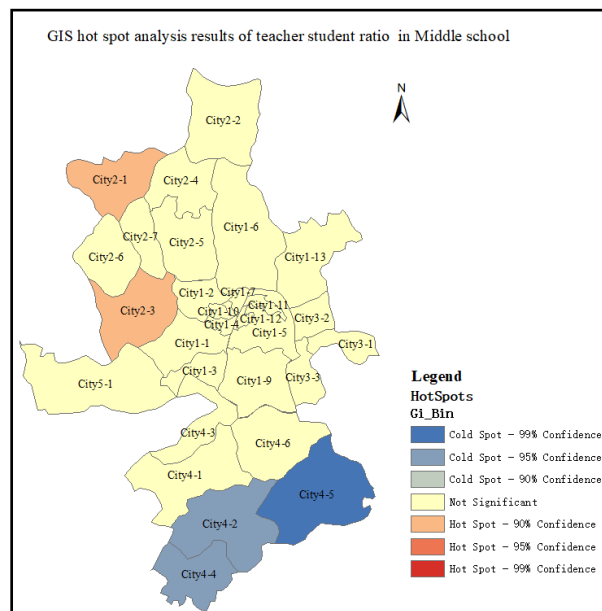

Fig 5: Hot Spots analysis results of teacher-student ratio of junior middle school in 30 districts

The hotspot analysis results of teacher-student ratio in high schools in Fig. 6 indicate that: the districts and counties under the jurisdiction of City1 exhibit strong spatial correlation, showingobvious high-value aggregation

characteristics, indicating that City1 has high index value of teacher-student ratio in overall in high school and enjoy more sufficient teacher allocation than other cities; while other districts and counties exhibit insignificant spatial correlation.
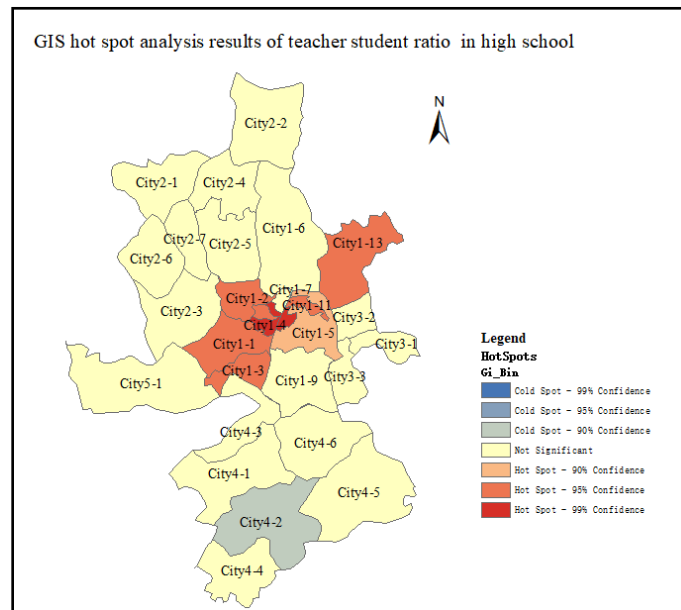

Fig 6: Hot Spots analysis results of teacher-student ratio of high school in 30 districts

V. Discussion

This paper mainly focuses on teacher-student ratio in compulsory education. Based on traditional data statistics and difference coefficient analysis, K-means clustering model and GIS hotspot analysis method are taken to analyze the educational resource clustering and spatial pattern and assess education equilibrium. The following conclusions are drawn:

(1) The five cities are all up to the standard in the allocation of teacher-student ratios in primary, junior middle and high schools. Difference coefficients are used to calculate the difference coefficients between the five cities and the districts under the jurisdiction of the five cities. The results indicate certain disequilibrium between different cities and between districts under the jurisdiction of the cities. Disequilibrium between different cities shows different significance level, which demands attention in light of the actual situation.

(2) A characteristic attribute vector is constructed for the teacher-student ratio of the three learning stages of the 30 districts under its jurisdiction.Through K-means clustering, it is found that the 30 districts show characteristics of category aggregation in the allocation of teachers, which can be aggregated into 3 categories;

(3) The teacher-student ratios of the three learning stages in the 30 districts under its jurisdiction also exhibit certain high-value aggregation and low-value aggregation characteristics in space. For example, high school in the provincial capital city exhibits obvious high-value aggregation. The three spatially adjacent areas of City2-1, City2-6, and City2-7 under the jurisdiction of city 2 also display high-value spatial aggregation in primary school, indicating sufficient teachers in this area as a whole

Compared with the existing research, the research methods and research ideas of this paperprovide a good exploration and supplement. For example, literature [6] and literature [9] mainly evaluate the allocation of educational resources from the perspective of numerical analysis. On the basis of statistical analysis of data, this paper uses K-means clustering to deduce category classification of educational resources input and regional clustering characteristics. In addition, using GIS to carry out education evaluation research, literature [12] conducts equilibrium evaluation mainly from the perspective of resource accessibility, while this paper conducts research

mainly from the perspective of spatial pattern of educational resource allocation and aggregation characteristics of high and low values. Hotspot analysis is made on 30 districts to deduce the spatial pattern of resource allocation and resource aggregation effect of teacher-student ratio. However, there are still some shortcomings in this research. For example, when using K-means and GIS hotspot analysis to evaluate educational resources, only one indicator of teacher-student ratio is selected for analysis. In the follow-up research, it is hoped that more evaluation indicators can be incorporated, such as teacher's professional title, teacher's academic qualifications, education funding and education informatization level, etc.

VI. Conclusions

Balanced development of compulsory education is an important way to achieve educational equity. Reasonably evaluating education equilibrium level and examining characteristics of educational resource input are important ways to narrow the gap in education levels between different schools, urban and rural areas, and different regions, thereby improving educational equity. This paper takes teacher-student ratio data in compulsory education stage as an example, selects basic data of 30 districts and counties in 5 cities,combines geographic space location, and uses difference coefficients, K-means clustering and GIS hotspot analysis method to analyze different teacher resource allocations and evaluate its equilibrium, explore the spatial aggregation mode of cold and hotspots in teacher resource allocation of different regions (cities, provinces/districts and counties), thus providing a new research idea for equilibrium evaluation of education and spatial pattern of resource allocation.

Acknowledgment

References

[1]   M. Tamayo,J. Rebolledo,A. Besoaín-Saldaña, "Monitoring inclusive education in Chile: Differences between urban and rural areas," International Journal of Educational Development,pp. 53110-116, 2017.

[2]   J. Kayiwa,K. Clarke,L. Knight,et al, "Effect of the good school toolkit on school staff mental health, sense of job satisfaction and perceptions of school climate: Secondary analysis of a cluster randomised trial," Preventive Medicine: An International Journal Devoted to Practice & Theory,pp. 101, 2017.

[3]   D Ayuso-Delpuerto, P Gutiérrez-Esteban, "Evaluation of Open Educational Resources through accessibility indicators," IV INTERNATIONAL VIRTUAL CONFERENCE ON EDUCATIONAL RESEARCH AND INNOVATION,2020.

[4]   Q. Xia, S.S. Shen, J.W. Wang, "Application of Regional Basic Education Resources and New Ideas of Construction Based on the Investigation of Jiangning District in Nanjing," Chinese audio-visual education, vol. 5, pp. 103-109+124, 2018.

[5]   R. Smyth, X. Qian, "Measuring regional inequality of education in China: widening coast-inland gap or widening rural-urban gap?," Journal of International Development,vol. 20, no. 2, pp. 132-144, 2010.

[6]   H.R. Zhang,Y.Q. Sheng, M. Luo, "Forty Years of the Balanced Development of Compulsory Education in China: Retrospection and Reflection," Journal of Southwest University (Social Sciences Edition),vol 45, no. 1, pp. 72-80+194, 2019.

[7]   E. Tomul, "Measuring regional inequality of education in Turkey: an evaluation by Gini index," Procedia - Social and Behavioral Sciences, vol. 1, no. 1, pp. 949-952, 2009.

[8]   H.M. Peng,Q. Lin,G.W. Wan, "Child population, economic development and regional inequality of education resources in China," Children and Youth Services Review,vol. 110,pp. 104819, 2020.

[9]   X.Q. Li, Q. Li, "Empirical Test of the Equilibrium of Compulsory Education in China from 2005 to

2015," Education Science,vol. 35, no. 3, pp. 67-74, 2019.

[10]   Y.L. Shen, X.Y. Qiao, "An analysis of the differences in the allocation of teacher resources in basic education in India," SHANGHAI JIAOYU KEYAN, vol. 3, pp. 50-52+72, 2012.

[11]   D.X. Zhou, "Status Analysis and Counter measures study on allocation of the resources of dual structure-urban teachers," Education&Economy, vol. 3, pp. 33-36, 2008.

[12]   X.Y. Zhang, "Analysis on the balance of educational resources and its spatial layout adjustment based on GIS," East ChinaNormalUniversity,2019.

[13]   R.G. Taylor, M.L. Vasu, J.F. Causby, "Integrated planning for school and community: The case of Johnston county," North Carolina. Interfaces,vol. 29, no. 1, pp. 67-89, 1999.

[14]   Y.Z. Wu, X.H. Zheng, L. Sheng, H.Y. You, "Exploring the Equity and Spatial Evidence of Educational Facilities in Hangzhou, China,"Social Indicators Research,vol. 151, pp. 1075–1096, 2020.

[15]   M.F. Guo, B.S. Hu, "Assessing the Spatial Layout Efficiency of basic Educational Resources by an Improved Coverage Model:A Case Study of Nanchang City, China," Journal of Geo-Information Science,vol. 21, no. 6, pp. 875-886, 2019.

[16]   H.Y. Liu, W. Chen, "A review of research on the distribution of basic education resources in China," PROGRESS IN GEOGRAPHY, vol. 36, no. 5, pp. 557-568, 2017.

[17]   F.F. Razi, "A hybrid DEA-based K-means and invasive weed optimization for facility location problem," Journal of Industrial Engineering International,vol. 15, pp. 499–511, 2019.

[18]   K. Wei-Chieh,Y.M. Lu,C Hao-Chi,et al., "GIS-Based Synthetic Measurement of Spatial Distribution and Hotspot Analysis on Aquarium in Taiwan," 2019 IEEE International Conference on Architecture, Construction, Environment and Hydraulics (ICACEH) pp. 12, 2019.

[19]   N. Shaheen,B. Raza,A.R. Shahid,H. Alquhayz, "A Novel Optimized Case-Based Reasoning Approach With K-Means Clustering and Genetic Algorithm for Predicting Multi-Class Workload Characterization in Autonomic Database and Data Warehouse System,"IEEE Access,vol 8, pp. 105713 – 105727, 2020.

[20]   C. Castro, Marcia, H. Burton, "Controlling the False Discovery Rate: A New Application to Account for Multiple and Dependent Test in Local Statistics of Spatial Association," Geographical Analysis, vol. 38, pp.180-208, 2006.