Volume 2, Issue 5, May 2012 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Analysis of Spatial Data Mining and Global Autocorrelation

M.R.Pavan Kumar
Department of CSE
Sree Vidyanikethan Engg College
Sivapavan.mr@gmail.com

K.S.Ranjith*
Department of CSE
Sree Vidyanikethan Engg College
ksranjith2000@gmail.com

B.Kiran Kumar
Department of CSE
Sree Vidyanikethan Engg College
bkiran1351@gmail.com

G.Mahesh Yadav
Department of CSE
Sree Vidyanikethan Engg College
golla.maheshyadav@gmail.com

Abstract— In this paper is to study of the spatial distribution of regional per capita agricultural total output value in Beijing rural regions of 2005 using exploratory spatial data analysis method of spatial data mining highlights the importance of spatial interactions and geographical location in regional issues. ESDA method including spatial weights matrix, Moran I index, Moran scatter plot and LISA is proved to be a powerful tool in this paper to reveal the characteristics of agricultural economy of each rural region in relation to those of its geographical environment.study the regional economic difference with the spatial data mining theories. townships in 2005 from the spatial interactive angel, and then reveal the spatial autocorrelation and spatial heterogeneity among townships. The results show that agricultural economy of Beijing townships has a strong spatial correlation generally, and there also exist spatial heterogeneity problems between local townships.

Keywords— spatial data mining; spatial analysis; regional economy**;** Global Autocorrelation

I. INTRODUCTION

The study of the spatial distribution of regional per capita agricultural total output value in Beijing rural regions of 2005 using exploratory spatial data analysis method of spatial data mining highlights the importance of spatial interactions and geographical location in regional issues. ESDA method including spatial weights matrix, Moran I index, Moran scatter plot and LISA is proved to be a powerful tool in this paper to reveal the characteristics of agricultural economy of each rural region in relation to those of its geographical environment.

First, ESDA reveals significant positive global spatial autocorrelation and very strong spatial cluster characteristics to regional rural economy of Beijing. Second, the analysis of Moran scatters plot and LISA diagram reveals that there exists spatial autocorrelation and spatial heterogeneity of per capita agricultural total output value between Beijing rural regions. The exist of spatial heterogeneity shows that global autocorrelation analysis indeed mask some local information and the local autocorrelation analysis give more help to recognize spatial outliers for more research. The relationships between spatial and non-spatial data from spatial database, as well as other data characters hidden in the database. Furthermore, the spatial data mining and knowledge discovery techniques aim at mining potential useful knowledge we have known nothing from existed database. Therefore, in recent years, the techniques of spatial data mining have been becoming a research hotspot of space information fields, and a great many important results have been achieved. The disparity of regional economy has always been the hot topic in regional economics field. Most of the traditional measure methods that have been used identical to investigate the disparity of regional economy, but the research indicates that it's a universal problem that there exists spatial disparity in regional economic development, patial effects, particularly spatial autocorrelation and spatial heterogeneity, must be taken into account when analysing convergence processes at regional scale . Therefore, in recent years, more and more researchers try to involve the spatial theory to the traditional methods of economic analysis and decision, and research economic activities and phenomenon on various scales from spatial thinking or spatial angle.

Exploratory spatial data analysis method (ESDA) of spatial data mining is a set of techniques aimed at describing and visualizing spatial distributions, at identifying atypical localizations or spatial outliers, at detecting patterns of spatial association, clusters or hot spots, and at suggesting spatial regimes or other forms of spatial heterogeneity. Through analysis of spatial autocorrelation, we can use the exploratory spatial data analysis method to reveal spatial dependency and spatial heterogeneity, and with further research, we can get a deep understanding of regional economic problems. In this paper, we take the rural areas of Beijing as example, and take the township as spatial scale research unit. With the correlation analysis of ESDA in spatial data mining, we describe and visualize the spatial patterns and characteristics of agricultural economy in Being rural areas.

II. spatial data analysis

In this paper, with exploratory spatial data analysis (ESDA) of spatial data mining, from the angle of spatial interaction, we get the spatial dependency and spatial heterogeneity through spatial autocorrelation analysis, and on this basis we investigate the space-time characteristics of regional economic disparity in Beijing rural areas.

ESDA is essentially the "data-driven" analysis method, with the spatial correlation measures as its core. The spatial autocorrelation is an important index to test the coincidence of value similarity with location similarity. We carry out the global statistics and the local statistics by computing spatial autocorrelation index. In this paper, we firstly normal transform research variables to meet the precondition of spatial autocorrelation analysis, then with the software GeoDa designed by Anselin to get spatial weight matrix, on this basis finally we measure global or local spatial autocorrelation using Moran's I or Local Moran I (LISA) index.

A. **Spatial Matrix**

Spatial weights matrix W is the precondition of exploratory spatial data analysis. The appropriate choice of the spatial weight matrix is one of the most difficult and controversial methodological issues in exploratory spatial data analysis and spatial econometrics.

The matrix W is defined as follows:

$$\begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1n} \\ W_{21} & W_{22} & \cdots & W_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ W_{n1} & W_{n2} & \cdots & W_{nn} \end{bmatrix}$$

Where n is the region numbers. If regions *i* and *j* is neighbor relationship, *Wij*=1; else *Wij*=0. Generally, we hold that a region has no neighbor relationship with itself, that is *Wii*=0. Given the characteristics of our sample of Beijing rural areas, as without sub district offices, the spatial weights matrix W used in this paper on the basis of distance rule, to avoid some errors arise from "islands" or "holes" problems.

B. **Spatial Autocorrelation**

When the same attribute of different observe objects in space present some regularity, but not random distribution, we can believe that there exists some spatial autocorrelation between observe objects. The measurement of global spatial autocorrelation is usually based on Moran's I statistic, this statistic is written in the following matrix form:

$$I = \frac{N\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{(\sum_i \sum_j w_{ij})\sum_i (x_i - \bar{x})^2}$$

Where *N* is the observe regions; *wij* is spatial weights matrix; *xi* is the value of region *i*; *x* is the mean value of all observation values. Moran's I varies between -1 and 1. On a given significant level, a value near 1 indicates that similar attributes are clustered, and a value near -1 indicates that dissimilar attributes are clustered. If a Moran's I is close to 0, it indicates a random pattern or absence of spatial autocorrelation.

C. **Local Spatial Autocorrelation**

The way to detect local spatial clusters but also to analyze local instability in the form of atypical localizations, spatial outliers, and spatial regimes is to use Moran scatter plots in conjunction with LISA as suggested by Anselin (1995). In the presence of global positive autocorrelation, Moran's I statistic may indeed mask regions that deviate from this global pattern

Anselin (1995) proposed a local Moran index or local indicator of spatial association (LISA) to capture local pockets of instability or local clusters. The local Moran index for a region i measures the association between a value at I and values of its nearby areas, defined as:

$$I_i = \frac{(x_i - \bar{x})}{s_x^2}\sum_j [w_{ij}(x_j - \bar{x})]$$

\

A positive *i I* means either a high value surrounded by high values (high - high) or a low value surrounded by low values (low - low). A negative *i I* means either a low value surrounded by high values (low - high) or a high value surrounded by low values (high - low).

III. RURAL REGIONAL ECONOMY

In this paper, we take the agricultural total output value per capita as index variable, and take the township as the basic analysis unit to research the rural regional economy spatial distribution of Beijing.

A. **spatial autocorrelation**

Firstly, the normal transformation was applied to research variable of the agricultural total output value per capita to fulfill the demand of global spatial autocorrelation test. With the formula 1, we continue to calculate the research variable's spatial correlation coefficient and the corresponding standardized statistics as follow table:

TABLE 1 MORAN'S I STATISTICS FOR RESEARCH VARIABLE

Variable	Moran's I	Standard deviation	Standardized value
per capita agricultural total output value	0.3997	0.0319	12.58

Note: The expected value for Moran's *I* statistic is constant: *E (I)* = -0.0054
All statistics are significant at $p = 0.001$

Table 1 displays the Moran's I statistic of the agricultural total output value per capita of 2005 for the 186 townships of Beijing rural region. Inference is based on the permutation approach with 1000 permutations. It appears that the agricultural total output value per capita is positively spatially autocorrelated since the statistic is significant with p=0.001. This result suggests that the distribution of per capita regional agricultural total output value is by nature clustered.

B. Local spatial autocorrelation

1) MORAN SCATTER PLOT.

In this paper, we adopted the GeoDa software designed by Anselin to calculate the Moran scatter plot of per capita regional agricultural total output value of 2005.

Figure 1 displays the Moran scatter plot for our sample: 186 townships of Beijing rural regions. It can be seen that most Beijing rural regions are characterized by positive spatial association. In 2005, 46.3% of Beijing rural regions exhibited association of similar values (22.6% in quadrant HH and 23.7% in quadrant LL). Furthermore, the Moran scatter plot can help to identify atypical regions. In 2005, 15
rural regions of Beijing displayed association of dissimilar values (11 in quadrant HL and 4 in quadrant LH). Some rural regions of chittoor District, kurnool District, Nellore District and Warngal District appear to be richer than their neighbors and are of the HL type. Only four LH rural regions are detected: An Ding ,HuaiRou, Liu Dian and WangXinZhuang .

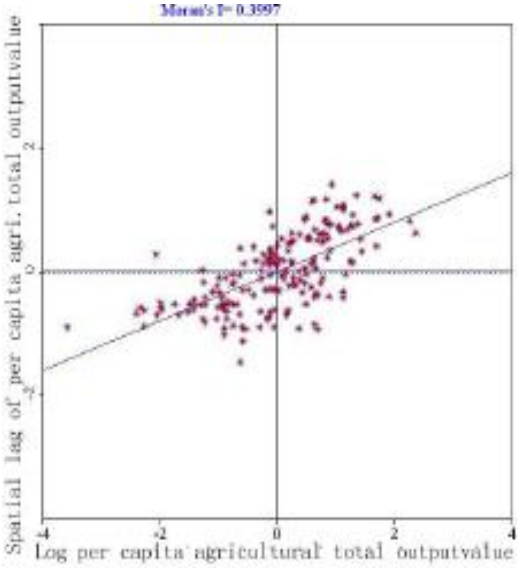

Figure 1. Moran Scatter Plot

2) Local Spatial Autocorrelation and LISA

LISA is a local spatial correlation index to measure the similar or dissimilar degree between research region and its neighbors. With GeoDa, we calculated local Moran I of per capita agricultural total output value of every township in Beijing, and plotted the LISA distribution diagram on the basis of z-test（ $p \leq 0.05$).

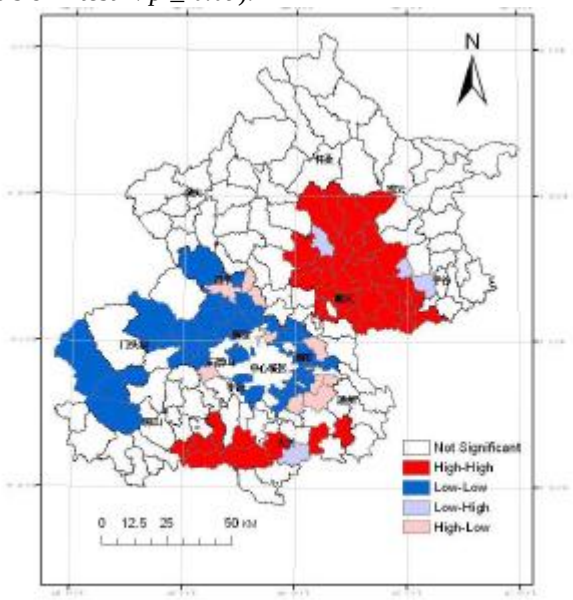

Figure 2. LISA Cluster Map

With the global spatial autocorrelation statistic, we can measure the spatial correlation degree, and then to capture the spatial correlation structure from the whole view. Since global spatial autocorrelation statistic yields a single result for the entire data set, it cannot discriminate between a spatial clustering of high values and a spatial clustering of low values in the case of a global positive spatial autocorrelation , and the global statistic may indeed mask regions that deviate from this global pattern. So we need to analyze the regional agricultural economy of Beijing from local scale. From the Moran scatter plot and LISA diagram, we concluded that: For the research variable of per capita agricultural total output value, such clusters are detected. For instance, the core regions and some western regions of Beijing are LL state. It means that the per capita agricultural total output values of these regions are all lower than others, at the same time, these regions cluster together. While the eastern regions of Beijing such as chittoor, Kurnool regions and some regions of southern of Beijing are HH state. From the calculated data, we know that 22.6% of Beijing rural regions are HH, and 23.7% are LL, that is there are nearly 50% of Beijing townships in the distribution status of agricultural total output value present cluster state. In addition, more insight into the scatter plot and LISA diagram, we can find that there exists spatial heterogeneity (HL or LH). Although the amount of these townships is of small proportion, is only 8%, the "hotspots" or "outliers" phenomenon should deserve more attention to research.

IV. Conclusion

The study of the spatial distribution of regional per capita agricultural total output value in Beijing rural regions of 2005 using exploratory spatial data analysis (ESDA) method of spatial data mining highlights the importance of spatial interactions and geographical location in regional issues. ESDA method including spatial weights matrix, Moran I index, Moran scatter plot and LISA is proved to be a powerful tool in this paper to reveal the characteristics of agricultural economy of each rural region in relation to those of its geographical environment.

First, ESDA reveals significant positive global spatial autocorrelation and very strong spatial cluster characteristics to regional rural economy of Beijing. Second, the analysis of Moran scatters plot and LISA diagram reveals that there exists spatial autocorrelation and spatial heterogeneity of per capita agricultural total output value between Beijing rural regions. The exist of spatial heterogeneity shows that global autocorrelation analysis indeed mask some local information and the local autocorrelation analysis give more help to recognize spatial outliers for more research.

REFERENCES

[1] Miller, Harvey J, Han Jiawei, Geographic Data Mining and Knowledge Discovery, Taylor & Francis, London, 2001.
[2] Di KaiChang, Spatial Data Mining and Knowledge Discovery, Wuhan University Press, Wuhan, 2001.
[3] Li Deren, Wang Shuliang, Li Deyi, "Theories and Technologies of Spatial Data Mining and Knowledge Discovery", Geomatics *and Information Science of Wuhan University*, 2002, pp. 221-233.
[4] Li Xiaojian, Qiao Jiajun, "Coun ty Level Econom ic D ispar it ies of Ch ina in the 1990s", ACTA GEOGRAPH ICA SINICA, 2001, pp. 136-145.
[5] Guo Qingwang, Jia Junxue, "The Factors Contribution to Regional Economic Convergence and Disparity in China", Finance & Trade Economics, 2006, pp. 11-17.
[6] Hu Liangmin, Miao Changhong, Qiao Jiajun, "A Study on the Divergence and Temporal-spatial Structure of Regional Economic Development in Henan Province", PROGRESS IN GEOGRAPY, 2002, pp. 268-274.
[7] Ou Xiangjun, Chen Xiuying, "Appraise of Integral Whole of Town and Country in China and Some Suggests for Public Policy", ECONOMIC GEOGRAPHY, 2004, pp. 338-343.
[8] Goodchild M, Anselin L, Applebaum R, "Towards a spatial integrated social Science", International Regional Science Review, 2000, pp. 139-159.
[9] Chen Fei, Guo Chaohui, "A Preliminary Study on Regional Economic Analysis and Decision-Making Based on GIS", HUMAN GEOGRAPHY, 2007, pp. 76-80.
[10] Meng Bin, et al, "Evaluation of Regional Disparity in China Based on Spatial Analysis", SCIENTIA GEOGRAPHICA SINICA, 2005, pp. 393-400.
[11] He Jiang, Zhang Xinzhi, "ESDA of the Regional Economic per capita
GDP in China", Statistics and Decision, 2006, pp. 72-74.
[12] Lu Feng, Xu Jianhua, "Exploratory Spatial Data Analysis of the Regional Economic Disparities in Chin", Journal of East China Normal University (Natural Science), 2007, pp. 45-51.
[13] Fan Xinsheng, Li Xiaojian, "On the Spatial Autocorrelation of Economic Growth Based on County Scale —A Case Study of Henan Province", Economic Survey, 2005, pp. 57-60.
[14] Chen Jianjian et al, "Analysis of Spatial and Temporal Characteristic of Regional Economic Disparity in Anhui Province since 1990s'", Journal of Anhui Agricultural, 2007, pp. 2633-2635.
[15] Anselin L. Interactive techniques and exploratory spatial data analysis. In: Longley P A, Goodchild M F, Maguire D J, et al. Geographical Information Systems, Principles, Technical Issues,
Management Issues and Applications. John Wiley & Sons, Inc, 1999. pp. 253-266.
[16] Cliff A D, Ord J K, Spatial Autocorrelation, Pion, London, 1973.
[17] Cliff A D, Ord J K. Spatial Processes: Models and Applications, Pion, London, 1981.
[18] Wang Yuanfei, He Honglin, Spatial Data Analysis Method, Science Press, Beijing, 2007.
[19] Anselin L, Spatial econometrics: methods and models, Kluwer Academic Publishers, Dordecht, 1998.
[20] Anselin L, Exploring Spatial Data with GeoDa: A Workbook, 2005.

[21] Julie Le Gallo, Cem Ertur, "Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980-1995", Regional Science, 2003, pp. 175-201
[22] Anselin L, "Local indicators of spatial association-LISA", Geographical Analysis, 1995, pp. 93-115

