Acta Technica 62 No. 1B/2017, 289–300

Mining urban active point circle based on spatio - temporal constraint data[1]

Rong Tao[2,3], Mengluo Ji[2]

Abstract. In order to detect the active points of cities and analyze the active circles in time and accurately, this paper proposes a method to detect urban active points and active circles based on the data of social network. A data preprocessing model based on discrete point is proposed to solve the problem of large data volume, discrete space-time constraint data storage and cluster analysis efficiency. Spatio-temporal constraint data were tested by spatial autocorrelation, which indicated that it had significant spatial clustering characteristics. This paper proposes an active clustering method based on spatio-temporal constraint data and explores the geography distribution of business factors to obtain active circle information. A city, for example, on the positioning network (www.dingwei.com) until September 30th, time and space constraints data test the active city point detection and active circle excavation test. The results show that there is a strong correlation between the active circle distribution and urban planning active circle based on spatio-temporal constraint data mining, which can be used to forecast the urban social economic development and regional economic planning.

Key words. Public-source geographic data, space-time constraint data, data mining, active-point detection, active-cycle distribution.

1. Introduction

As one of the driving forces of urban economic development, urban active circle is an important part of urban comprehensive competitiveness. The active circle dynamic measurement is an important basis for guiding the economic layout of the city. It plays a very important role in bringing into full play the social benefits and overall functions of the active circle, promoting the process of urbanization and promoting the sustained development of the national economy. At home and abroad active circle research mainly has macro angle and micro angle two aspects

[1]This work was support by the Henan province foundation and advanced technology research project in 2015 (No.152300410115).

[2]College of Computer and Information Engineering, Luoyang Institute of Science and Technology, 471023, Luoyang, China

[3]Corresponding author

[1]. The former extension of the measurement range is too wide, will make active lap determination doping many other factors, which focuses on the micro level of the enterprise active lap analysis, are not from the city perspective to the city active lap determination. In addition, commonly used active circle measurement method is generally used questionnaire survey method, takes more time and effort, limited scope of the investigation, affecting the active circle determination of accuracy and comprehensiveness [2].

The crowd sourcing geographic data is the open geographic data collected by the public and provided to the public [3–5]. The representative geographical data includes GPS trajectory data, map data compiled by users' collaborative annotation, and points of interest (POIs) for various social networking sites such as Twitter, Facebook, and the street (www. Dingwei. Com), etc. [6]. Compared with the traditional geographic information data, the public geographic data from the non-professional public has the characteristics of large data volume, good potentiality, abundant information and low cost, and has become the active research area of international geographic information science in recent years [7–10]. Space-time constraint data is a kind of data which has spatial, temporal and social attribute information produced by the intelligent terminal with GPS. It records the life trajectory and reflects people's daily life behavior, an important source of geographical data. Spatiotemporal constraints data concentrated in the city, and to the public to sign the points of interest as the main form. To positioning network, for example, since May 13, 2010 positioning network officially launched since the location network registered users to 20 % per week to maintain the rapid growth rate, as of September 2011, the number of positioning network users has more than 1.2 million. As more than 70 % of registered users will sign-up information and social platform binding, so the average sign-in will have 400 audiences. Location network every 5 s (based on 24 h basis) to update a user check-in information, including a wealth of location information, semantic information and behavioral information. Therefore, the space-time constraint data obtained from the positioning network not only rich in data, but also good potential, from the side to reflect the city's economic and cultural distribution situation. Based on spatial and temporal constraint data of positioning network, this paper presents a method to detect urban active point and active circle based on all-source spatial and temporal constraints data. Through data preprocessing, exploratory spatial analysis and spatial clustering analysis, constraint data high-value clustering active point.

2. Exploratory clustering and analysis of spatio-temporal constraint data

Spatio-temporal constraint data is a discrete GIS point object with spatial coordinates and user attributes. First, the discrete log-in data is lattice processed, and the large data volume and discrete sign-in point are transformed into spatial continuity and adjacency. Grid data that reflects the density of check-in events. Secondly, spatial correlation of measured data is measured by exploratory spatial data analysis (ESDA), and its spatial structure and global distribution pattern are measured to determine the best mode of active point detection and active circle clustering. Thirdly, clustering analysis was used to identify the locations of active, cold and spatial anomalies with statistical significance. Finally, by measuring the spatial distribution of the clustering results, the spatial characteristics of clustering geographic elements are obtained, including the range of the active circle, the central trend and the direction of development. The concrete algorithm flow is shown in Fig. 1.

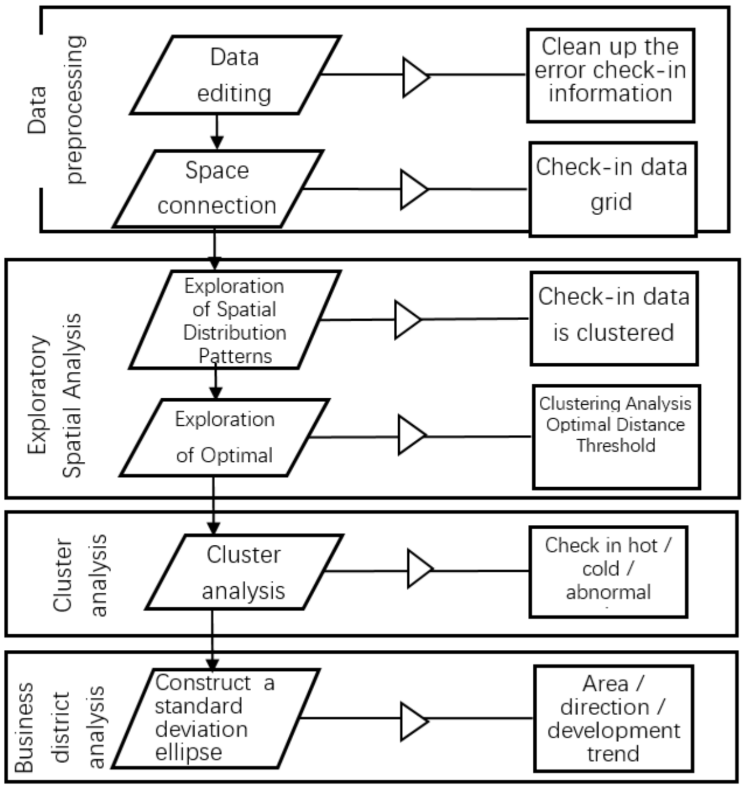

Fig. 1. Flowchart of active point detection and active circle mining algorithm based on spatio-temporal constraint data

2.1. Spatio-temporal constraint data grid processing

Spatiotemporal constraint data is a large number of discrete feature points, which does not have obvious spatial continuity and adjacency in space, and is not conducive to the exploration of spatial data analysis method to measure its spatial distribution pattern. In order to make space-time constraint data reflect both spatial continuity and proximity, and to preserve the characteristics of sign-in times and the statistical properties of key objects, this paper chooses the area covered by sign-in data as the research area, where $G\,(W)$ is the grid G, and G is the space between the point-to-sign and data-key attributes of the grid-containing data, and the corresponding grid attributes are mapped to the corresponding grid attributes [11, 12] of the sign

$$\left.\begin{aligned}
&G\left(W\right)=\textstyle\sum_{i=1}^{n}Np_i\times\sigma p_i,\quad \text{among them }\ P_i\subset\left(\{p\}\bigcap G\right),\\
&G\left(T\right)=T_k,\ \ \text{among them }\textstyle\sum\sigma p,T_k=\\
&\max\left\{\textstyle\sum\sigma p,T_1,...,\sum\sigma p,T_i,...,\sum\sigma p,T_s\right\}.
\end{aligned}\right\}\tag{1}$$

Here, $G\left(T\right)$ represents the area type of the grid G, n represents the number of check-in points in the grid G, Np_i represents the total number of check-in points of the ith grid in the grid G, σp_i represents the weighting level of the check-in point, $\sum\sigma p, T_i$ represents the sum of the weights of all check-in points belonging to the Kth class in the grid G. Figure 2 shows the algorithmic flow of gridding of discrete sign-on data grids.

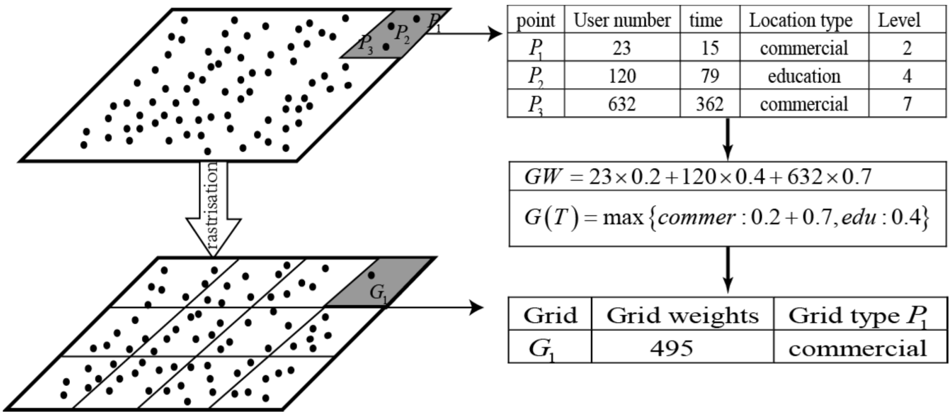

Fig. 2. Schematic diagram of grid-based processing of temporal-spatial constraint data

From Fig. 2, the discrete sign-in data is transformed into the grid-signed data set with the sign-in frequency as gray scale, which not only simplifies the discrete point data, but also preserves the time-space characteristic and thematic attribute characteristic of the signed data, satisfies the exploratory spatial analysis and data Mining requirements.

2.2. Spatio-temporal constraint data space autocorrelation test and analysis

Spatial autocorrelation is an important form of spatial dependency and a prerequisite for exploratory spatial data analysis (ESDA). In this paper, global spatial autocorrelation of space-time constraint data is studied by global Moran's statistical method, and Ripley's K-statistic method is used to explore the spatial distribution pattern with the strongest sign-in feature to provide a basis for sign-in data mining.

Given a set of spatio-temporal constraint data and their check-in frequency, Global Moran's I-statistics is evaluated as a clustering pattern according to (2), where n is the number of check-in points, z_i being the sign-in position. Here

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} z_i z_j}{S_0 \sum_{i=1}^{n} z_i^2}\,, \tag{2}$$

where z_i is the check-in frequency of the deviation $\left(x_i - \bar{X}\right)$ from the mean, $w_{i,j}$ is the spatial join matrix of the check-in position, S_0 is the sum of all spatial weights, given by the formula $S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}$ denoting molecular normalization by variance, the index value being in the range –1.0 to +1.0. Its positive value indicates that the sign-in frequency of the sign-in position has a positive correlation, and a negative value indicates a negative correlation, indicating that the spatial distribution of the spatial object is not correlated. This can be calculated in accordance with equation (3) to check that the value of the sign frequency is statistically significant—EOT.

$$z_I = \frac{I - E[I]}{\sqrt{V[I]}}\,. \tag{3}$$

Here,

$$E[I] = \frac{-1}{n-1},\ V[I] = E\left[I^2\right] - E^2[I]\,.$$

Figure 3 shows the spatial distribution pattern of spatial-temporal constraint data based on Global Moran's I statistical computation. The score value for the test is 8.003898 times the standard deviation, much larger than 2.58, indicating that the null hypothesis probability value is 0, consistent with the requirement of 99 % confidence value (probability likelihood value). The global spatial autocorrelation of the spatial pattern of constraint data is in accordance with the statistical characteristics of the typical clustering model, which can be used for cluster analysis of active points and active circles.

2.3. Geographical distribution measures of check pointing data clustering active points

The high-value active point detected by local autocorrelation clustering can be regarded as the center of active circle, and its range, center change and direction should be further defined. In this paper, the active circle is studied by measuring the geographical distribution of commercial clustering active points, Specific steps are as follows:

2.3.1. Standard deviation elliptical structure. The standard deviation ellipses are constructed with the clustering active points as the center and the sign-in positions and their associated attribute values (frequency of sign-in). The ellipse center is the weighted average center of the feature in the clustering area, and the ellipse length and short axis of the feature distribution are defined by the standard distance

Fig. 3. Distribution of urban active points

in the x and y directions.

$$SDE_x = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n}}\,,$$

$$SDE_y = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \bar{Y}\right)^2}{n}}\,, \tag{4}$$

Here SDE_x and SDE_y are the long and short axes of the standard deviation ellipse.

2.3.2. Calculation of active circle range based on standard deviation ellipse. Calculation of active circle range based on standard deviation ellipse. With the standard deviation ellipse length and the short axis as the spatial distribution of the active circle, the major axis is the main trend of the central trend of the active cycle.

2.3.3. Determination of active circle direction. The proportion of the long and short axes represents the flatness of the active circle distribution, and the rotational azimuth of the standard deviation ellipse is the development trend of the active circle. The rotation angle of standard deviation ellipse is calculated according to the formula

$$\tan\theta = \frac{\Lambda + B}{C}\,. \tag{5}$$

Here,

$$\Lambda = \left(\sum_{i=1}^{n} \tilde{x}_i^2 - \sum_{i=1}^{n} \tilde{y}_i^2\right), \quad B = \sqrt{\left(\sum_{i=1}^{n} \tilde{x}_i^2 - \sum_{i=1}^{n} \tilde{y}_i^2\right)^2 + 4\left(\sum_{i=1}^{n} \tilde{x}_i \tilde{y}_i\right)^2},$$

$$C = 2\sum_{i=1}^{m} \tilde{x}_i \tilde{y}_i \,.$$

Here, $\tilde{x}_i$ and $\tilde{y}_i$ are the standard deviations of the clustered feature points with respect to the center point of the ellipse.

The local standard deviation ellipse is constructed according to the geographical distribution measurement method and the range, the central trend and the trend of each active circle are calculated. The result shows the spatial-temporal constraint data of a certain city.

3. Experiment and discussion

3.1. Analysis of detection results of active city based on spatio-temporal constraint data

According to the attribute information of the high-value sign-in point of each active area, this paper divides the active area of the clustering into commercial active points (such as Jianghan Road), educational active points (such as a city university), tourism (Such as the East Lake area), traffic class active points (such as light rail station), living class active point (such as South Lake district) and other types of active points (such as restaurants, libraries) six categories, the spatial distribution shown in Fig. 3.

As shown in Fig. 3, the commercial activity points are clustered in the spatial distribution, which reflects the active circle distribution of a city. The number of commercial activity points aggregated by each active circle reflects the popularity of the active circle. Figures 3 and 4 in the mouth and the mouth of the street, the mouth of the street area of commercial high active point number, indicating a relatively large population of street population, the economy has maintained a relatively active state in the Simingkou area. On the contrary, the region's commercial activity is only 3 points, indicating that the door area population flow is not, the economy is not active and there are signs of recession, and the two regional economic development in line. Figure 4 shows the statistical distribution of different types of active points. Of the 172 high-value sign-in elements, there are 90 commercial activity points, 43 educational activity points, 12 tourism activity points, 10 traffic activity points, class activity points 3, 14 other types of active points. It can be seen that the proportion of business activity points in space-time constraint data is the most, which validates the rationality of using spatio-temporal constraint data to analyze active circle.

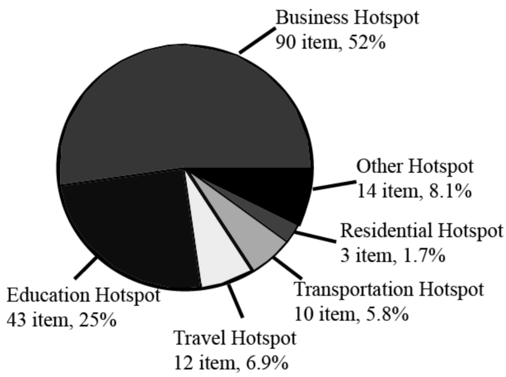

Fig. 4. Classification map of urban active points

3.2. Analysis of urban active circle of a city based on spatio-temporal constraint data

The results show that a city has formed commercial and business sections with Wuchang, Hankou and Hanyang as the main pattern, and dozens of large and medium - sized active laps with a certain scale. Both the traditional sense of the old active circle, such as Jianghan Road, Zhongshan Road, Wangjiawan, Zhongjiacun, Xudong, the door of the door, fruit lake, active circle in the South, there are emerging in recent years active circle, World, street mouth, Lu Xiangguang Valley active circle, but also includes some active in the construction and development of the circle, such as water chestnut Lake, Dunkou zone active circle. For each of the 90 commercial economic activity points, the standard deviation ellipses are used to compute the range, trend and direction of each active circle, and the spatial distribution of each active circle in a city based on sign-in data is obtained (Fig. 5). Through the spatial superposition analysis, the statistical results of the number of sign-in points, the number of sign-in times, the number of registered users, and the average number of sign-in times of users in each active circle are obtained.

It can be seen from Table 2, based on spatial and temporal constraints of data analysis of a city active circle with a city three towns active circle actual distribution. According to the statistics of active circles in three cities, the statistical distribution characteristics of Wuchang, Hankou and Hanyang in terms of business activity points, sign-in number and number of users coincide with the regional functional characteristics and population distribution characteristics. The number of active users in Wuchang has exceeded that of Hankou, and the average number of its users is the highest among the three towns. It can be seen that Wuchang, as the center of traditional science, culture and politics, has a tendency to overtake Hankow as a traditional economic center is relying on the East Lake High-tech Development Zone, the formation of policy-oriented Lvxiang Optics Valley active circle and the integration of Asian Trade Shopping Center, the IT port of the IT market in Guangzhou port active lap effect. In contrast, in a city of Hanyang Economic and Technological Development Zone, commercial scale in the initial stage of development, although the average number of attendants and Wuchang quite, but the active circle of commercial outlets rather small. Compared with the Hankou business model, the Wuchang active circle is generally discrete in geographical distribution, showing a clear regional distribution of active laps. Hankou area, the traditional Jianghan Road active circle, Wuhan-Guangzhou World Trade active circle is still playing a pillar of the role of mid-stream, a new city active circle, water chestnut Lake (Wanda) active circle also plays an increasingly important role.

Table 2. Active lap information statistics table

Active circle	Business active points	Attendance	User number	The average number of attendance
Central and South active circles	126	2206	1615	1.3569
Division door active circle	104	3019	2374	1.2717
Street mouth active circle	658	15412	9179	1.6791
Fruit Lake active circle	67	1038	672	1.5446
Lu Xiangguang active circle	407	14771	9264	1.5948
Xudong active circle	95	3463	2069	1.6738
Marshland active circle	71	2851	1704	1.6731
A city active circle	131	4955	3592	1.3795
Jianghan Road active circle	421	4790	3293	1.4546
Northwest Lake active circle	258	4133	2882	1.4341
Wuhan-Guangzhou World Trade active circle	300	7781	5050	1.5408
Lake water chestnut active circle	48	3900	2621	1.4880
Zhongjiacun active circle	170	2184	1445	1.5114
Wangjiawan active circle	66	1377	957	1.4389
Dunkou active circle	34	869	515	1.6874
Wuchang	1457	39912	25173	1.5855
Hankou	1229	28410	19142	1.4842
Hanyang	270	4430	2917	1.5187

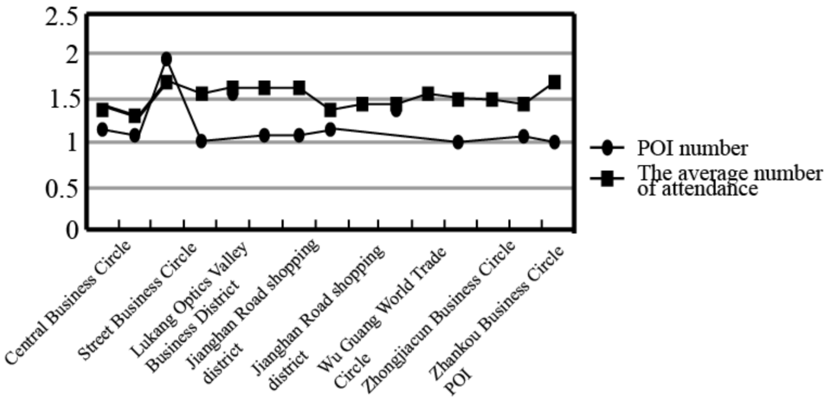

Fig. 5. Active circle POI and the average number of attendance

It can be seen from Fig. 5 that in all active circles of a city, the number of commercial outlets and the average number of attendance in the street lap are all at the highest level in the whole city, and it is the largest IT computer in the city of Guangbaihuo, New World, Market, but also with a city traffic is currently the most congested point of the same, we can see that the source of space-time constraint data for active lap active point detection is very effective.

4. Conclusion

The emergence of all-source geographic data has provided a new data source for spatial data updating, and provided a new research direction for spatial data mining. In this paper, we propose a method to detect and analyze the active points of the active points of interest accumulated in the city for a long time. The clustering analysis based on the number of the endorsements of the points of interest and the construction of the standard deviation ellipse can accurately determine the city Active range and distribution of the active circle, from the meso-level use of spatial analysis and mining and other means of urban active and active circle detection and analysis. Compared with the traditional method of active circle measurement and analysis, this method has the characteristics of objectivity, real-time and high accuracy. It is shown that the spatial and temporal constraint data has obvious clustering characteristics by the test results overlaid with a city administrative map. This paper is based on the analysis of the data from the spatial and temporal constraints to find the distribution of the active circle in a city is consistent with the objective facts and more detailed. This result reflects the high correlation between daily life behavior and the distribution of commercial economy in a certain city area. It provides a new method for monitoring the distribution and development trend of urban commercial circle. It also provides a new method for city planning and administration Decision-making provides a more intuitive reference. It is necessary to further study the automatic active point detection method based on spatio-temporal constraint data. At the same time, the active circle analysis can be used to get the active circle dynamic changes, such as the change of the active circle and the change of the active circle range, real-time monitoring the active circle's rise, and decline.

References

[1] Z. X. Liao, W. C. Peng: *Clustering spatial data with a geographic constraint: Exploring local search.* Knowledge and Information Systems *31* (2012), No. 1, 153–170.

[2] L. J. Yang, H. L. Zhu, J. P. Wu: *Market area analysis in retailing based on GIS.* Remote Sensing technology and application *18* (2003), No. 3, 144–148.

[3] J. Giles: *Wikipedia Rival Calls in the Experts.* Nature *443* (2006), No. 7111, paper 493.

[4] J. Howe: *The rise of crowdsourcing.* Wired magazine *14* (2006), No. 6, 1–4.

[5] C. Heipke: *Crowdsourcing geospatial data.* ISPRS Journal of Photogrammetry and Remote Sensing *65* (2010), No. 6, 550–557.

[6] B. T. Haworth: *Implications of volunteered geographic information for disaster management and GIScience: A more complex world of volunteered geography.* Annals of the American Association of Geographers (2017) 1–15.

[7] B. J. Hecht,M. Stephens: *A tale of cities: Urban biases in volunteered geographic information.* International Conference on Weblogs and Social Media (ICWSM), 1–4 June 2014, Ann Arbor, Michigan, USA, Publisher AAAI press (2014), 197–205.

[8] P. Neis, D. Zielstra: *Recent developments and future trends in volunteered geographic information research: The case of opensreetmap.* Future Internet *6* (2014), No. 1, 76–106.

[9] M. F. Goodchild, J. A. Glennon: *Crowdsourcing geographic information for disaster response: A research frontier.* International Journal of Digital Earth *3* (2010), No. 3, 231–241.

[10] S. Fritz, I. McCallum, C. Schill, C. Perger, R. Grillmayer, F. Achard, F. Kraxner, M. Obersteiner: *Geo-Wiki.org: The use of crowdsourcing to improve global land cover.* Remote Sensing *1*, (2009), No. 13, 345–354.

[11] A. M. Lu, C. M. Li, Z. J. Lin, W. Z. Shi: *Spatial continuous surface model of population density.* Acta Geodaetica Et Cartographica Sinica *32* (2003), No. 4, 344–348.

[12] Z. Halim, M. M. Gul, N. ul Hassan, R. Baig, S. Ur Rehman, F. Naz: *Malicious users' circle detection in social network based on spatio-temporal co-occurrence.* International Conference on Computer Networks and Information Technology, 11–13 July 2011, Abbottabad, Pakistan, IEEE Conference Publications (2011), 35–39.

Received June 29 7, 2017