A BUFFER ANALYSIS BASED ON CO-LOCATION ALGORITHM

Guoqing Zhou[1], Shengxin Huang[1], Haoyu Wang[1, *], Rongting Zhang[2], Qingyang Wang[1], Hongjun Sha[1],Xiaofan Liu[1], Qiuyu Pan[1]

[1] Guangxi Key Laboratory of Spatial Information and Geomatics, Guilin University of Technology, No. 12 Jian'gan Road, Guilin, Guangxi 541004, China - (gzhou, haoyu)@glut.edu.cn
[2] School of Precision Instrument and Opto-Electronic Engineering, Tianjin University, Tianjin 300072, China

Commission III, WG III/1

KEY WORDS: Buffer analysis, Co-location, Space Analysis, Space mining, Co-location analysis

ABSTRACT:

Buffer analysis is a common tool of spatial analysis, which deals with the problem of proximity in GIS. Buffer analysis researches the relationship between the center object and other objects around a certain distance. Buffer analysis can make the complicated problem be more scientifically and visually, and provide valuable information for users. Over the past decades, people have done a lot of researches on buffer analysis. Along with the constantly improvement of spatial analysis accuracy needed by people, people hope that the results of spatial analysis can be more exactly express the actual situation. Due to the influence of some certain factors, the impact scope and contact range of a geographic elements on the surrounding objects are uncertain. As all we know, each object has its own characteristics and changing rules in the nature. They are both independent and relative to each other. However, almost all the generational algorithms of existing buffer analysis are based on fixed buffer distance, which do not consider the co-location relationship among instances. Consequently, it is a waste of resource to retrieve the useless information, and useful information is ignored.

1. INTRODUCTION

Due to the influence of certain factors, the range of influence and contact range of a geographical element on the surrounding objects is uncertain. However, existing buffer analysis generation algorithms are based on a fixed buffer distance regardless of the co-location between the instances. Zhou *et al*. (2010) combined GIS and spatial data mining to solve pavement repair decision problems. Zhou *et al.* (2012) proposed a method called co-location decision tree, CL-DT, which uses Euclidean distance to determine whether there is a spatial cooperative juxtaposition between two instances. Zhang (2016) improved the CL-DT algorithm and proposed an MVU-based CL-DT algorithm to perform calculations on instances to better express the distance between instances. This paper proposes a buffer analysis method based on the co-location algorithm. Co-location mining patterns are used to guide the generation of point buffers, line buffers, and polygon buffers.

2. GENERALIZED BUFFER SPATIAL ANALYSIS

2.1 Co-location algorithm introduction

Huang *et al*. (2004) first presented the general framework of mining spatial co-location patterns. Where after, in order to make clear how to apply mining co-location rules in spatial data analysis, spatial pattern classification, and spatial geographic knowledge discovery, Huang and her research team made further research and exploration, such as Huang *et al*. (2008), Xiao *et al*. (2008) and Wang *et al*. (2003). Yoo et al. (2009) proposed partial-join algorithm. Yoo et al. (2006) proposed join-less algorithm and N-most prevalent collected event. Verhein *et al*. (2007) proposed complex spatial co-location patterns.
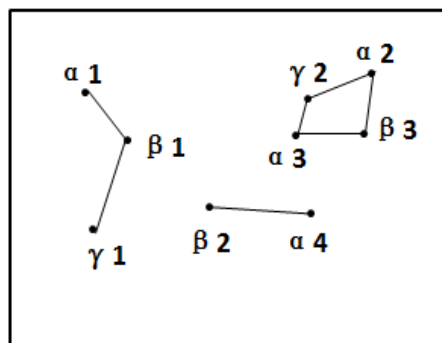

Figure 1. Co-location mode diagram

It has several important definitions:

Definition 1: Spatial objects are different features in the spatial range. It is a collection of different object categories. Set a spatial object set to P= {*p1*，*p2*，⋯，*pj*}. The data amount of the spatial data set *Si* is *i*. *i*∈ [1, *j*], Si contains all instances of the spatial object *Pj*. In Figure 1. The figure contains three spatial objects *α, β, γ*.

Definition 2: The R-proximity relationship describes the proximity relationship between spatial instances. Its value can be represented by *R*. The spatial object satisfies the *R* value to indicate that the spatial relationship is satisfied.

Definition 3: The set of instances *K* satisfies R-proximity relationship, then it is called a row instance of co-location *Ω*.In Figure 1. Co-location *Ω*= {*α, β, γ*}. {*α1, β1, γ1*} is a row instance of co-location *Ω*. {*α, β, γ*} contains {*γ2, α3, β3*} and {*γ2, α2, β3*}. They are row instance. The collection of all row instances is called a table instance of co-location *Ω*.

* Corresponding author: Haoyu Wang,Email:haoyu@glut.edu.cn.

Definition 4: Participation rate (*PR*) is a specific value obtained from instances of instances that do not repeat. It can be expressed by

$$PR(\Omega, p_j) = \frac{\pi_{\Omega i}(table\_instance}{table\_instance} \tag{1}$$

Where π is the projection operation of the relation. For example, $\pi_{\Omega i}$ Table_instance ({Ω}) is a projection of a table instance of co-location Ω on p_j.

Definition 5: Participation (PI) is the minimum value of *PR*. When *PR* is greater than or equal to the threshold. The co-location model is a frequent co-location model . In Figure 1. There are 4 instances of object α, 3 instances of object β, and 2 instances of object γ. Assuming co-location $\Omega=\{\alpha,\gamma\}$, The table instance of Table_instance Ω is {$\alpha1$, $\gamma1$}, {$\alpha3$, $\gamma2$}, {$\alpha2$, $\gamma2$}. PR (Ω, α) = 3/4, *PR* (Ω, γ) = 2/2. *PI* = *PR* (Ω, α) = 3/4.

2.2 Co-Location-based Buffer Analysis Algorithm

2.2.1. Point buffer based on co-location mining： The buffer is a rule circle with the target point as the center and a certain distance as the buffer radius. Variable K as the buffer radius, There is a co-location relationship between K and the boundary point. The co-location based point buffering algorithm is implemented as follows:
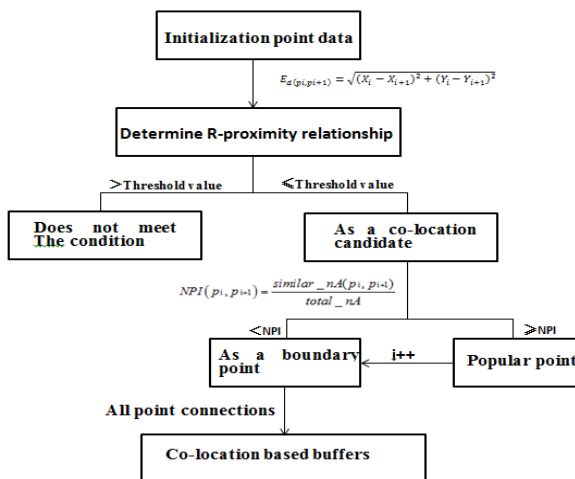

Figure 2. Irregular point buffer implementation process

Spatial coordinates are important parameters for spatial analysis. So let the space object be *Pi* (*Xi, Yi*). The distance between two spatial objects is

$$E_d(p_i, p_{i+1}) = \sqrt{(X_i - X_{i+1})^2 + (Y_i - Y_{i+1})^2} \tag{2}$$

Determine the distance between two points. Set NPI to determine whether candidate points satisfy the co-location relationship.

$$NPI(p_i, p_{i+1}) = \frac{similar\_nA(p_i, p_{i+1})}{total\_nA} \tag{3}$$

Similar_nA is the distance from the center point to the neighboring point. Total_nA is the radius of the traditional buffered data. NPI takes a larger value as the boundary point.

2.2.2. Line buffers based on co-location mining： In a traditional line buffer curve, the object is considered an axis. On both sides of the axis, the parallel lines of the axis consist of a constant deviation. For the end point and the inflection point, a convex arc method or an angle dichotomy method is used to generate a circular arc that intersects the parallel lines to form a closed area. Taking into account the actual situation, according to the conditions of the boundary points, use a cubic spline interpolation method to form a closed area. Making the buffer distance is no longer a constant, but a variable that is constrained by the co-location rules, and the constructed buffer is more in line with the actual situation.
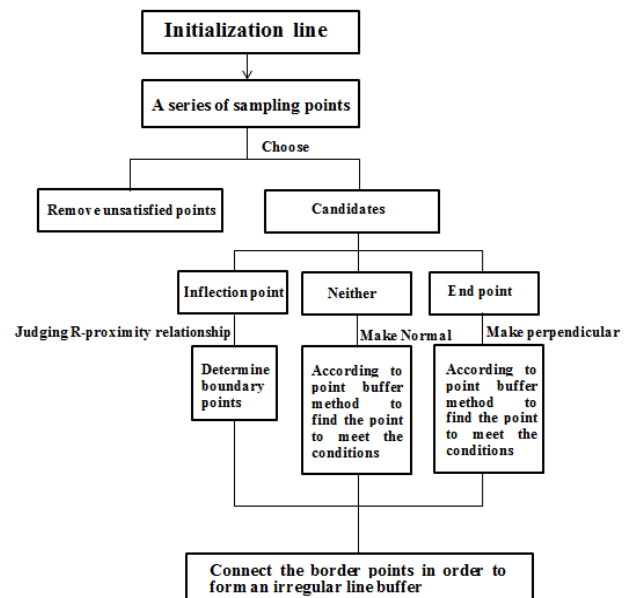

Figure 3. Irregular line buffer implementation process

2.23. Polygon buffer based on co-location mining： The polygon buffer based on co-location mode mining is established in the same manner as the point and line.
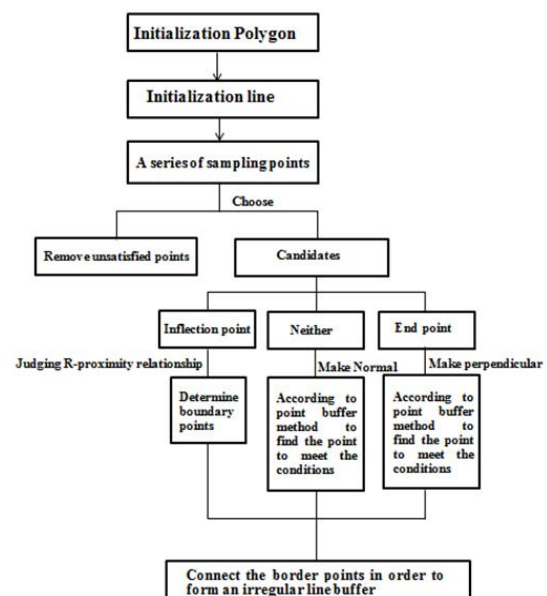

Figure 4. Irregular polygon buffer implementation process

3. EXPERIMENT AND DISCUSSION

This experiment is used 2015 Shenzhen Baoan District map data, Including point, line, surface data sets, analysis tool is ArcGIS10. The following three different data sets for irregular buffer generation, as well as compared the irregular buffer with the traditional buffer. Compared with the traditional curve buffer generation algorithm, the proposed co-location algorithm takes into account the co-location relationships between instances and instances .so that the buffer distance is no longer a constant, but a variable constrained with co-location positioning regulation , the buffer building more suit with the actual situation.

3.1 Irregular point buffer generation

The R-proximity relationship threshold is set to 11000m. The threshold is the buffer radius and constitutes a traditional point buffer. The points in the buffer are all candidates for co-location for easy screening. The distance correlation between the X and Y attribute values of the center point and the point within the buffer is selected as a reference. According to the participation rate (PR) and the participation index (PI), the distance from each candidate point to the center point is compared with the threshold. The difference between them is more frequent as a co-location candidate, and the difference between them is a common point, it is, a boundary point. Each common candidate point is connected in turn to form an irregular point buffer based on the co-location algorithm.
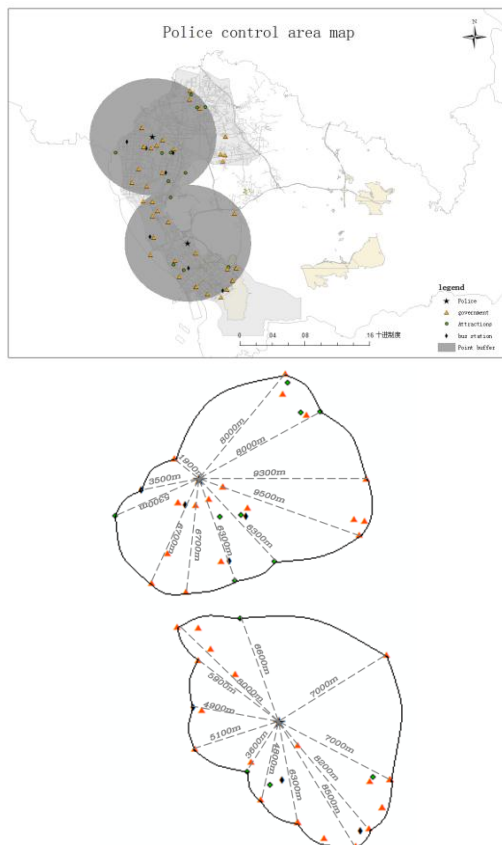

Figure 5. Point buffer comparison chart (a. Traditional buffer; b.Irregular buffer)

The selected case study object has two center points that would cause overlap areas when generating traditional buffers. In the overlap section covers a few co-location candidates, It is not

clear at a glance which of the traffic police detachments the adjoining points belong to. This will lead to unnecessary troubles caused by accidents in real life. The irregular buffers formed based on co-location can be accurate to the specific instance of jurisdiction.

3.2 Irregular line buffer generation

Taking these two curves as the central axis, a threshold value of 800 m is set according to the R-proximity relationship, and the threshold is also used as a buffer radius to establish a traditional buffer zone. The co-location relationship between the center axis and the adjacent points in the region is studied to establish an irregular buffer. Point within the area as a co-location candidate point. Because the river is line feature and the co-location relationship is to study the location relationship between an instance and another instance, thus used ArcMap to deal with the line feature, the line features are transformed into a lot feature points, and the feature points are filtered to select the sample points , Because this experiment relies on examples to analyzed and build up the speed of water supply in rivers and establish the optimal position of water supply pipelines, the sampling points are selected according to the Euclidean distance formula, and their   selected are based the adjacent point is perpendicular to the center Axis. As for the experiment, 25 sampling points are selected, and then the sampling points are taken as the central point. The same as the method of selecting the same points, the common boundary points are selected one by one from the surrounding candidate points, and the boundary points are sequentially connected to make up the co-location Buffer based on line element irregularity.
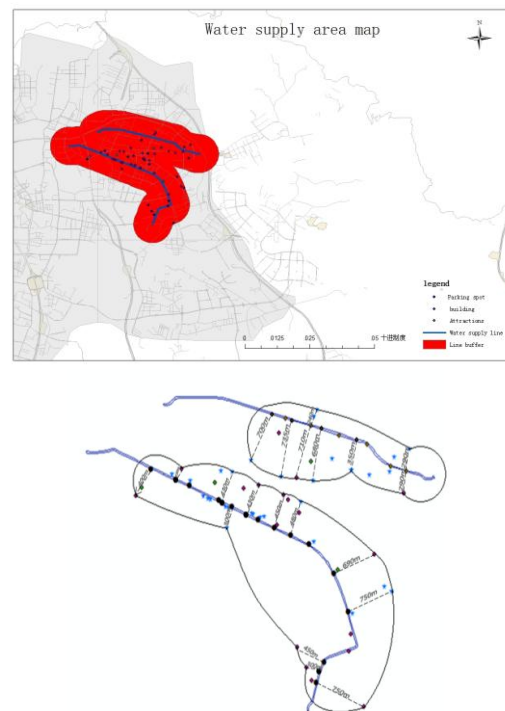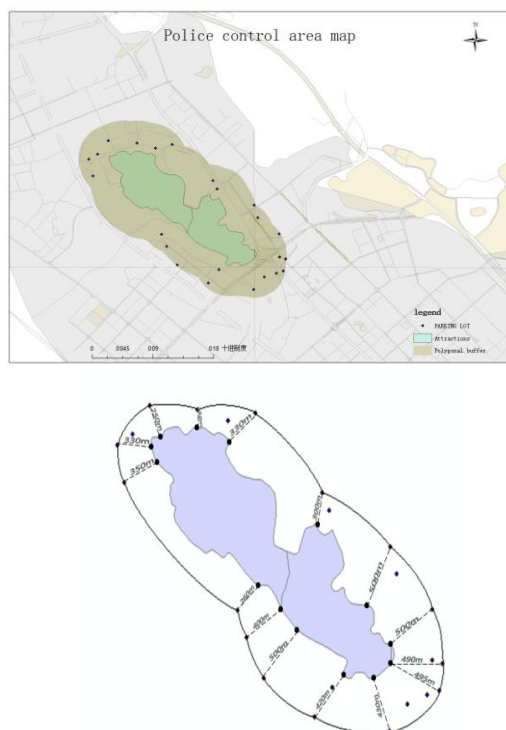

Figure 6. Line buffer comparison chart (a. Traditional buffer; b.Irregular buffer)

Because the constraints of the actual conditions, the water supply of the river should be covered and accurate to various points. The traditional buffer area is too wide to cover the area of influence of the river water supply area, but it could not be specific to the water supply object and the establishment of the

water supply network cannot be determined. The irregular buffer based on co-location can establish the specific relationship between the river and the object point according to the correlation. Easy to maintain and repair.

3.3 Irregular surface buffer generation

Tourism attractions can be seen as an irregular polygon in the map. According to the R-proximity relation, a threshold of 500m is selected as a buffer radius to establish a traditional buffer. Parking spots in this area are all regarded as co-location candidates. The boundaries of tourist attractions can be seen as a curve, so you can refer to the selection of line elements, select the sampling points, and then these sampling points can be the tourist import and export attractions. According to the selection of point features, selected the optimal parking position, make up an irregular surface buffer based on co-location.

Figure 7. Polygonal buffer comparison chart (a. Traditional buffer; b.Irregular buffer)

The traditional buffer only gives a rough range of parking but does not pinpoint the exact parking location. However, the distribution of parking lots is rather scattered, which may lead to traffic jams in the peak tourist season. In addition, the irregular buffers based on the formation of co-location, we can distribute the parking resources reasonable, and choose the best parking location, convent for the management of work staff.

4. CONCLUSION

The buffer analysis based on the co-location algorithm proposed in this paper is based on the spatial co-location pattern mining .Combines the ideas of buffer analysis in spatial analysis. The buffer is no longer a constant, and the buffer becomes a specific variable based on the relevance of the spatial object. This is more convenient for research and application in actual production. It can make complex problems more scientific and intuitive, and provide users valuable information. It also

eliminates useless information to reduce resource waste and maximize the resource utilization.

REFERENCES

Huang, Y., Shekhar, S. and Xiong, H., 2004. Discovering colocation patterns from spatial data sets: a general approach. *IEEE Transactions on Knowledge & Data Engineering*, 16(12), pp.1472-1485.

Huang Y., Zhang P. and Zhang C., 2008. On the relationships between clustering and spatial co-location pattern mining. *International Journal on Artificial Intelligence Tools*, 17(1), pp.55-70.

Jin, S. Y. and Shekhar, S., 2006. A joinless approach for mining spatial colocation patterns. *IEEE Transactions on Knowledge & Data Engineering*, 18(10), pp.1323-1337.

Jin, S. Y. and Bow, M., 2009. Finding N-Most Prevalent Colocated Event Sets. *International Conference on Data Warehousing and Knowledge Discovery* Springer-Verlag, 5691, pp.415-427.

Verhein, F. and Al-Naymat, G., 2007. Fast Mining of Complex Spatial Co-location Patterns Using GLIMIT. *IEEE International Conference on Data Mining Workshops, 2007. ICDM Workshops* IEEE, pp.679-684.

Wang F., Zhang Z. and Machemehl R. B., 2003. Decision-making problem for managing pavement maintenance and rehabilitation projects. *Journal of the Transportation Research Board*, 1853(1), pp.21-28.

Xiao X., Xie X., Luo Q. and Ma W., 2008. Density based co-location pattern discovery. *In Proceedings of the 16th ACM Sigspatial International Conference on Advances in Geographic Information Systems* DBLP, pp.1-10.

Zhou, G. and Wang L., 2010. GIS and Data Mining to enhance pavement rehabilitation decision-making. *Journal of Transportation Engineering*, 136(4), pp.332-341.

Zhou, G. and Wang L., 2012. Co-location decision tree for enhancing decision-making of pavement maintenance and rehabilitation. *Transportation Research Part C*, 21(1), pp.287–305.

Zhou, G., Zhang, R. and Zhang, D., 2016. Manifold learning co-location decision tree for remotely sensed imagery classification. *Remote Sensing*, 8(10), 855.