

# Strategy Model Based on Geographic Profiling

Na Wang\*, Xinshe Qi, Guo Li and Jing Li

College of Information and Communication, National University of Defense Technology

Xi'an 710106, China

\*Corresponding author

**Abstract**—In order to generate a geographical profile, based on the spatial distribution theory and the probability distance theory, two strategy models are established. At the same time, we also analyze the complexity and the accuracy of these models.

**Keywords**—spatial distribution strategy; probability distance strategy; crime prediction

## I. INTRODUCTION

As well known, every crime action must be in some particular geographical time-space. Thereby, the crime action always manifests some geographical regularity influenced by geographical environment.

Spatial analysis has long been a valuable tool used within the criminal investigative process<sup>[1, 2]</sup>. This is especially true for serial offence cases, where criminologists apply geographic profiling to model the offender mobility and the crime distribution patterns in order to estimate a criminal's likely residence<sup>[3, 4]</sup>. Yet, traditional analytical methodologies have avoided the utilization of the functional distance measures when modeling an offender's journey-to-crime within an anisotropic landscape.

One of the focuses of any police investigation is the crime scene and its evidentiary contents. What is often overlooked, however, is a geographic perspective on the actions preceding the offence, the spatial behavior that led up to the crime scene. For any violent crime occurring, there must have been an intersection of victim and offender in both time and place. How did this happen? What were the antecedents? What do the spatial elements of the crime tell us about the offender's actions and hunting patterns?

## II. SPATIAL DISTRIBUTION STRATEGY

### A. Model of Spatial Distribution Strategy

In two-dimensional Euclidean space, the center of minimum distance of a given points set is the location, at which the sum of the distances to all other points is the smallest. These coordinates are derived through the iterative method

$$W(\bar{x}, \bar{y}) = \sum_{i=1}^n dist((x_i, y_i), (\bar{x}, \bar{y})) \quad (1)$$

where  $dist()$  is the Euclidean distance between the point  $(x_i, y_i)$  and the chosen point  $(\bar{x}, \bar{y})$ . Obtaining a precise value for this

complex strategy is more difficult than other spatial distribution strategy. The problem of minimizing  $W$  is not a trivial task, which has been studied for a long history. The function  $W$  is convex, so it has a unique minimum in the convex hull of the given points set. Weiszfeld developed an iterative algorithm, which can approximate the estimated home location through the iterative method as follows:

$$\begin{aligned} x^{(k+1)} &= \frac{\sum_{i=1}^n x_i / dist((x_i, y_i), (x^{(k)}, y^{(k)}))}{\sum_{i=1}^n 1 / dist((x_i, y_i), (x^{(k)}, y^{(k)}))}; \\ y^{(k+1)} &= \frac{\sum_{i=1}^n y_i / dist((x_i, y_i), (x^{(k)}, y^{(k)}))}{\sum_{i=1}^n 1 / dist((x_i, y_i), (x^{(k)}, y^{(k)}))} \end{aligned} \quad (2)$$

where the initial value  $(x^{(0)}, y^{(0)})$  may be taken as the centroid of the given points set<sup>[5]</sup>. Kuhn proved that the method actually converges to the center of minimum distance, but no precise estimate is known on how long it takes for this method to converge, as a function of the number of input points. Given  $(x^{(k)}, y^{(k)})$ , the computing algorithm of  $x^{(k+1)}$  and  $y^{(k+1)}$  involves the inverse of the distances between  $(x^{(k)}, y^{(k)})$  and the  $n$  given points. Then, the sum of  $n$  numbers computed in the previous step is computed, which is followed by the computation of  $x^{(k+1)}$  and  $y^{(k+1)}$ .

### B. Complexity Analysis of Spatial Distribution Strategy

In this section, we provide an example of a complexity analysis for a spatial distribution strategy<sup>[6-7]</sup>. In all cases, the computational complexity of implementing such strategy is at least linear with the number of crime locations. In each case,  $T(n)$  denotes the computational complexity of the particular method.

Given  $n$  points, the center of the circle is calculated as the mid-point of the two furthest points in the sequence of crime locations. We can compute the distance between  $(x_i, y_i)$  and  $(x_j, y_j)$  using the formula  $\sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$  with two subtractions, two multiplications, and a square root operation. If  $R$  is the cost of taking the square root of a number, then

$T \leq 4 + R + 3$ . The total cost of this method can be estimated as, therefore

$$T(n) \leq \left\{ (n-1) + (R+8) \times \left(\frac{n}{2}\right) + 4 \right\} \quad (3)$$

Although in general  $T(n) \geq n$  in this situation, any program computing the center of the circle of a set of  $n$  points will need to compute  $\left(\frac{n}{2}\right)$  distances. Hence, a more accurate lower bound on this method's complexity can be estimated as

$$T(n) \geq \left(\frac{n}{2}\right).$$

### III. PROBABILITY DISTANCE STRATEGY

#### A. Model of Negative Exponential

The negative exponential strategy assumes that, the likelihood of an offender living at a particular location is the highest near an offender's crime site location and decreases with increasing distance. As the equation  $f(x) = ae^{-bx}$  ( $a > 0$  and  $b > 0$ ) show that, the decline is at an exponential rate, dropping quickly near an offender's crime location, as it approaches zero likelihood. This prediction is represented by one of a family of distance decay functions, where  $a$  is a coefficient (equal to 10 as default) used to provide an indication of the maximum likelihood of finding a home, and  $b$  is an exponent (equal to 1 as default) that determines the slope of the function,  $x$  is the distance between each reference location and each crime location, and the cost of evaluating  $f$  is the likelihood of an offender living at a particular location.

#### B. Model of Truncated Negative Exponential

The truncated negative exponential strategy is a spline function consisting of the linear strategy and the negative exponential strategy<sup>[8]</sup>.

$$f(x) = \begin{cases} bx & \text{if } x \leq x_0 \\ bx_0 e^{-c(x-x_0)} & \text{if } x > x_0 \end{cases} \quad (4)$$

This is a complex function consisting of two distinct decay equations: linear and exponential. For locations in close approximation to the residences, a positive linear function can be defined as follows:

$$Pct_i = A + Bd_i \quad \text{for } 0 \leq d_{ij} \leq \text{Max } d_{ij} \quad (5)$$

Starting at zero (the haven) and increasing to a peak distance,  $\text{Max } d$ . Next, the function follows a negatively signed exponential function, declining quickly as the distance increases.

$$Pct_i = A \cdot e^{-cd_i} \quad \text{for } d_{ij} > \text{Max } d_{ij} \quad (6)$$

For points near the crime site locations, a positive version of the linear strategy can be used to represent the notion that offenders have a low likelihood of living directly on a crime location but an increased likelihood of living in the surrounding areas. At a predefined point of highest likelihood, the offender's "safe distance" away from the crime locations, the linear strategy is replaced by the negative exponential strategy to represent a declining likelihood of locating the offender's home with increasing distance. The total cost of calculating the truncated negative exponential is the addition of the complexities associated with the two strategies.

### IV. CONCLUSION

As mentioned before, probability distance strategy assumes that the crime site locations represent an offender's activity space, and that this space contains that offender's residence.

Given  $n$  input points (i.e., crime locations), a grid is laid on top of them. We let  $(u_1, v_1), (u_2, v_2), \dots, (u_m, v_m)$  be the centers of the various cells defined by such a grid. For each center,  $C_i = (u_i, v_i)$ , we can then define the Euclidean distance,  $dist_{ij}$  from  $C_i$  of the  $j$ th crime location,  $(x_j, y_j)$  ( $n$  values defined for each center). Then, a given probability density function  $f$  can be evaluated, and the numerical value associated to  $C_i$  is:

$$l(C_i) = \sum_{j=1}^n f(dist_{i,j})$$

The predicted home location is then set to be the center  $C$ , that maximizes function  $l$ . In other words, a numerical score can be assigned to a number of points in the region where the offences were committed, allowing us to identify the point with the highest probability score. For each of the probability distance strategies, we can use a common formula to measure complexity. Calling  $T_f$  the time to evaluate  $f$ , and assuming that  $m$  centers are defined, each of the methods in this section had an approximate complexity of at most  $m \times n \times T_f$  computational steps. Clearly, if  $m$  is much smaller than  $n$ , these methods can be simpler than many of the spatial distribution strategy. On the other hand, if  $m$  is bigger than  $n$ , then the complexity of any of these probability distance strategies is at least  $n^2$ .

### REFERENCES

- [1] B. Snook, M. Zito, C. Bennell, and P. J. Taylor, "On the complexity and accuracy of geographic profiling strategies," *Journal of Quantitative Criminology*, vol. 21, no. 1, pp. 1-26, 2005.
- [2] D. Brown, J. Dalton, and H. Hoyle, "Spatial forecast methods for terrorist events in urban environments," *Proceedings of International Conference on Intelligence and Security Informatics*, pp. 426-435. Tucson, USA, 2004.
- [3] H. Liu, and D. Brown, "Criminal incident prediction using a point-pattern-based density model," *International Journal of Forecasting*, vol. 19, no. 4, pp. 603-622, 2003.
- [4] M. Laukkonen, and P. Santtila, "Predicting the residential location of a serial commercial robber," *Forensic Science International*, vol. 157, no. 1, pp. 71-82, 2006.
- [5] N. Levine, "CrimeStat: A spatial statistical program for the analysis of crime incidents," S. Shekhar, H. Xiong, and X. Zhou, (eds) "Encyclopedia of GIS," Springer, Cham. pp. 381-388, 2017.

- [6] G. O. Mohler, and M. B. Short, "Geographic profiling from kinetic models of criminal behavior," *SIAM Journal on Applied Mathematics*, vol. 72, no. 1, pp. 163-180, 2012.
- [7] J. H. Ratcliffe, "A temporal constraint theory to explain opportunity-based spatial offending patterns," *Journal of Research in Crime and Delinquency*, vol. 43, pp. 3, pp. 261-291, 2006.
- [8] Y. F. Xue, and D. E. Brown, "Spatial analysis with preference specification of latent decision makers for criminal event prediction," *Decision Support Systems*, vol. 41, no. 3, pp. 560-573, 2004.