Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, 11-14 July 2010

A Novel Serial Crime Prediction Model Based on Bayesian Learning Theory

Renjie Liao[1], Xueyao Wang[1], Lun Li[2] and Zengchang Qin[1*]

[1]Department of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China
[2]Department of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China
*E-MAIL: zengchang.qin@gmail.com

Abstract

How to build affective mathematical models to understand the behaviors of serial crimes is an interesting research field in public security. Several theories have been proposed to handle this problem [1]-[4]. In this paper, we introduce a novel serial crime prediction model using Bayesian learning theory. There are many potential factors affecting a serial offender's selection of the next crime site, we mainly studied the factors related to geographic information. For each factor, by using a discrete distance decay function which derives from the classical crime prediction theory "Journey to Crime", we create a geographic profilewhich is a probability distribution of being the next crime site on given geographical locations. The final prediction is made by combining all geographic profiles weighted by effect functions which can be adjusted adaptively based on Bayesian learning theory. By testing the model on a crime dataset of a serial crime happened in Gansu, China, we can successfully capture the offender's intentions and locate the neighborhood of the next crime scene.

Keywords:

Crime Prediction, Geographic Profiling, Mixture of Gaussian Distribution, Bayesian Learning Theory, Kernel Function, Hausdorff Distance

1 Introduction

With the rapid development of the advanced information technologies, using new techniques to aid the crime investigating is of great importance, especially investigating the serial crime which is considerably cruel and complex.

Experts have studied the serial crime for many years. Based on the detailed data of serial crimes, they have get many credible theories about serial offenders. *Centrography* [1], *journey to crime* [2], *routine activity theory* [3] and *circle theory* [4] are four most significant theories in analyzing and predicting serial crimes.

The widely used method in these theories to forecast next location the criminal possibly offenses in a serial crime is geographical profiling. Geographical profiling is an investigative technique that analyzes the spatial pattern of a related series of crime locations in order to predict the location of the offenders residence [5][6]. Some researchers develop this theory to predict the probable future criminal positions of serial crimes. Stile and Brown [7] present a methodology for combining known geographical profiling methodologies with spatial event prediction in their work.

2 Crime Site Prediction Model

2.1 Factors

Considering the event of selecting the next crime site in a serial crime, it is obvious that various factors affect the serial offender. Based on the understanding of a large amount of crime data and research reports like [8][9], we classify these factors into two categories according to that whether they have the typically geographical distributions:

The factors related to the geographical distribution:

- Characteristics of the victims: gender, age, jobs, race and etc.
- Characteristics of the crime sites: private and public regions.
  - private regions: private residence, apartments etc.
  - public regions: school, subway, bus stop, hospital, square, park etc.

The factors unrelated to the geographical distribution:

978-1-4244-6525-5/10/$26.00 ©2010 IEEE

- Characteristics of modus operandi.
- Characteristics of criminal psychology

Based on the experience of practical criminal investigation, we concludes that it is difficult to get the cues unrelated to geographical factors at the crime scene, for example, the psychological characteristics of the offender. And even we get the evidence, it is difficult to quantify them for predicting the location of next crime. These cues can only supply some undirected help to local law enforcement offices. So, in our article, we only consider the factors related to the geographical distribution which will help us build geographic profile for crime site prediction.

2.2 Geographic Profile of a Specific Factor

- **Distance Decay Function:**

To describe the geographic profile of each factor, we introduce the distance decay function from [9] as below.

$$f(x) = \begin{cases} 0 & \|x-c\| \in (0, R_z) \\ \lambda & \|x-c\| \in (R_z, R) \\ \lambda e^{-\beta(\|x-c\|-R)} & \|x-c\| \in (R, +\infty) \end{cases} \tag{1}$$

where x is a location coordinate vector, λ is a normalization factor, c is a location coordinate vector of the center which indicates the residence of the offender in the *journey to crime* theory, R and R_z are empirical parameters. This function shows that the likelihood of a location being where an offender is based decays with the distance from the center. Here, we estimate the parameter c in (1) with the location coordinate vector of the center of the circle which is the smallest enclosing circle of known crime sites. However, for a specific factor, such as school, its geographic distribution is discrete, like Fig1.

Therefore, we can not use the distance decay function forementioned.

- **Discrete Distance Decay Function:**

To improve it, we use the mixture of Gaussian distribution to approximate the discrete geographic distribution of the factor and weight each Gaussian distribution by the distance decay function.

We assume that there are k discrete probable locations of next crime only considering one specific factor. For every location μ_i, we can use Gaussian model to calculate its probability distribution in the region of the whole town with the same value of σ. The equation (2) shows how to get this probability distribution.

$$g(x, \mu_i) = e^{-\frac{(\|x-\mu_i\|)^2}{2\sigma^2}} \tag{2}$$

Then, we weight all the locations' probability distributions with the consideration of distance decay function and add them. So, we can get the discrete distance decay function as below.

$$P(x) = \delta \sum_{i=1}^{n} \alpha_i g(x, \mu_i) \tag{3}$$

In the function (3), δ is the normalization factor. And α_i satisfies the distance decay function (1).

In fact, the discrete distance decay function is continuous within the location coordinate. However, we call it discrete because it adds the discrete geographic distribution of a specific factor. Through the discrete distance decay function, we can get the geographic profile of a specific factor.

3 Dynamic Prediction Model based on Bayesian Learning Method

Having got all these factors' geographic profiles, we can combine them to do the prediction. Here, we establish a dynamic prediction model which will learn from the history crime site series and reduce the prediction error by Byaesian learning method.

3.1 A Measure of Effect

We assume that there are a total of m factors. To combine all these factors' geographic profiles, we must first know how these factors affect a serial offender to make a decision about selecting next crime site. Therefore we build a effect function $u(i)$ to quantitatively study factors' effect on selecting crime site.

Before the first offense, we have no prior knowledge about how much each factor weight in the selecting decision. So it is reasonable to assume that every factor's effect is the same i.e.

$$u_1(i) = \frac{1}{m} \qquad i \in (1, 2, ..., m) \tag{4}$$

3.2 Bayesian Learning Method

For a specific offender, the hypothesis above may not stand. Therefore, we need to adjust the effect function dynamically based on the dynamical known crime data, especially data of crime sites. Here, we bring forward a adaptive adjustment algorithm and put it in a Bayesian learning framework.

- **Bayesian Learning Framework:**

Suppose that we've already got some information about the first n ($n \geq 1$) crime sites, we can set up the area's geographical profile based on discrete distance decay function

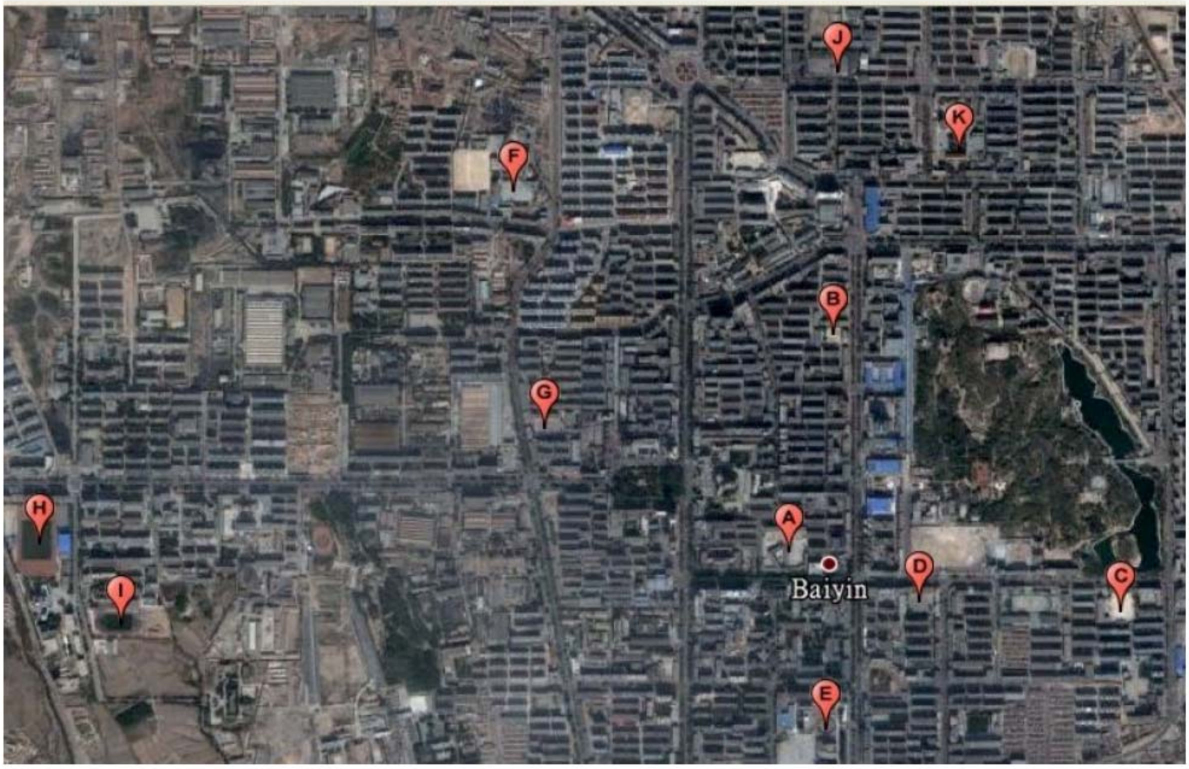

Figure 1. The Geographical Distribution of School.

with consideration of only one specific factor. For example, when we are only considering the subway factor, we can get the subway's geographical profile by taking the center of the smallest enclosing circle of first n crime sites. After getting each factor's geographic profile, we can combine them with the effect function like below

$$P(x) = \sum_{i=1}^{m} u_n(i) P_i(x) \tag{5}$$

$P_i(x)$ stands for the geographic profile with the consideration of the ith factor and $P(x)$ is the final geographic profile which can give the prediction. Convert it to the Bayesian form:

$$P(x|D_n) = \sum_{i=1}^{m} u_n(i) P_i(x|D_n) \tag{6}$$

where $P(x|D_n)$ indicates the dynamically changing final geographic profile on the premise that we have known the nth crime data.

For the geographic profile with the consideration of the ith factor $P(x|D_n)$, we can calculate it by Bayes theory:

$$P_i(x|D_n) = \frac{P_i(D_n|x)\,P_i(x)}{\rho} \tag{7}$$

where ρ is a normalization factor. Substitute it into (6):

$$P(x|D_n) = \sum_{i=1}^{m} \frac{u_n(i) P_i(D_n|x)}{\rho} P_i(x) \tag{8}$$

Then, the form of adjusting effect function is:

$$u_{n+1}(i) = \frac{u_n(i) P_i(D_n|x)}{\rho} = \frac{P_i(D_n|x)}{\rho} u_n(i) \tag{9}$$

- **Implement Adaptive Adjustment by Kernel Function:**

To reach the adaptive adjustment, we should compute function (9). Thinking that, if the predicted area given by considering a specific factor deviates much with the real crime site, it is most likely that this factor has less effect to the specific offender. Based on this premise, we use kernel function to implement adjustment as below:

$$\frac{P_i\left(D_n|x\right)}{\rho} = \begin{cases} \mathrm{K}\left(\mathrm{H}\left(X, y_{n+1}\right)\right) & \mathrm{H}\left(X, y_{n+1}\right) > r \\ \varepsilon & \mathrm{H}\left(X, y_{n+1}\right) \leq r \end{cases} \tag{10}$$

where $\mathrm{K}(\mathrm{H}(X, y_{n+1}))$ is the kernel function. Here we can use Gaussian kernel function for implementation. r is a empirical parameter which indicates the scope of the prediction area. The $\mathrm{H}(X, y_{n+1})$ represents the Hausdorff distance between the real crime site y_{n+1} and the prediction region X. The reason to define the distance as Hausdorff distance is that the prediction region X is a set of points. The form of Hausdorff distance is:

$$H(X, y_{n+1}) = \max\left[h(X, y_{n+1}), h(y_{n+1}, X)\right] \tag{11}$$

where, $h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|$.

The first equation of function (10) indicates that the real crime site is not included in the prediction region considering the ith factor only. The other one shows the real site is in the prediction region.

Assuming that there are t factors when considered respectively making the real crime site out of the prediction region, we adjust them as below:

$$u_{n+1}(i) = u_n(i)\mathrm{K}(\mathrm{H}(X, y_{n+1})) \tag{12}$$

Then, we can adjust the rest $(m - t)$ factors which make the the real crime site included in the prediction region as in (10):

$$u_{n+1}(i) = u_n(i)\varepsilon \tag{13}$$

In this equation, ε is the normalization factor. We use it to normalize the m factors' effect function.

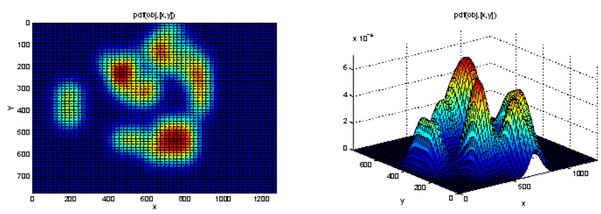

Figure 2. The Geographical Profiling of Apartment Factor.

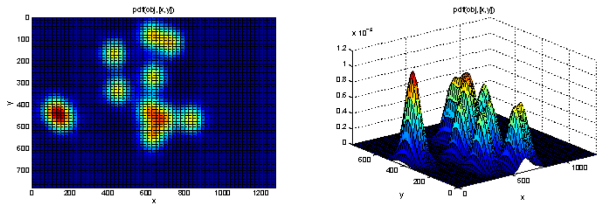

Figure 3. The Geographical Profiling of School Factor.

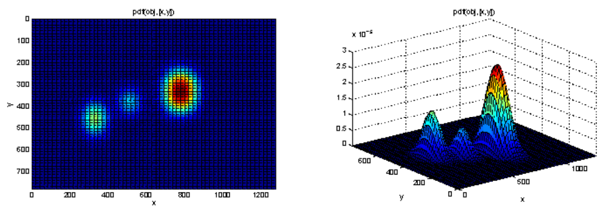

Figure 4. The Geographical Profiling of Park Factor.

4 Simulation

To test our model, we have investigate a lot of serial crime cases from various media resources, such as Internet, crime reports, newspapers etc. Since most serial crime data is kept in local law enforcement offices, we can only get limited detailed data of those notable serial crime cases. Through our filtration, we kept several serial crime cases whose time span is limited to several months and crime sites concentrate in cities and towns.

The result below is based on a case[10] which occurred in Baiyin city, Gansu Province, China. The offender committed for 16 years, killed 9 women and in 1998. He committed 4 consecutive crimes in 11 months. We choose typical factors which are related to geographical distribution as many as possible. However, due to the lack of data about characteristics of victim in Baiyin city of the year 1998, we only get factors of crime sites' characteristics. According to the principle that the more the number of distribution, the better the factor, we select 5 factors below: apartment, school, hospital, park and hotel.

We do the simulation as follow steps:

(1) We download the map of Baiyin city from Google.

(2) Make the man-lived parts of the map gridded into many rectangles with the same size which will help a lot to

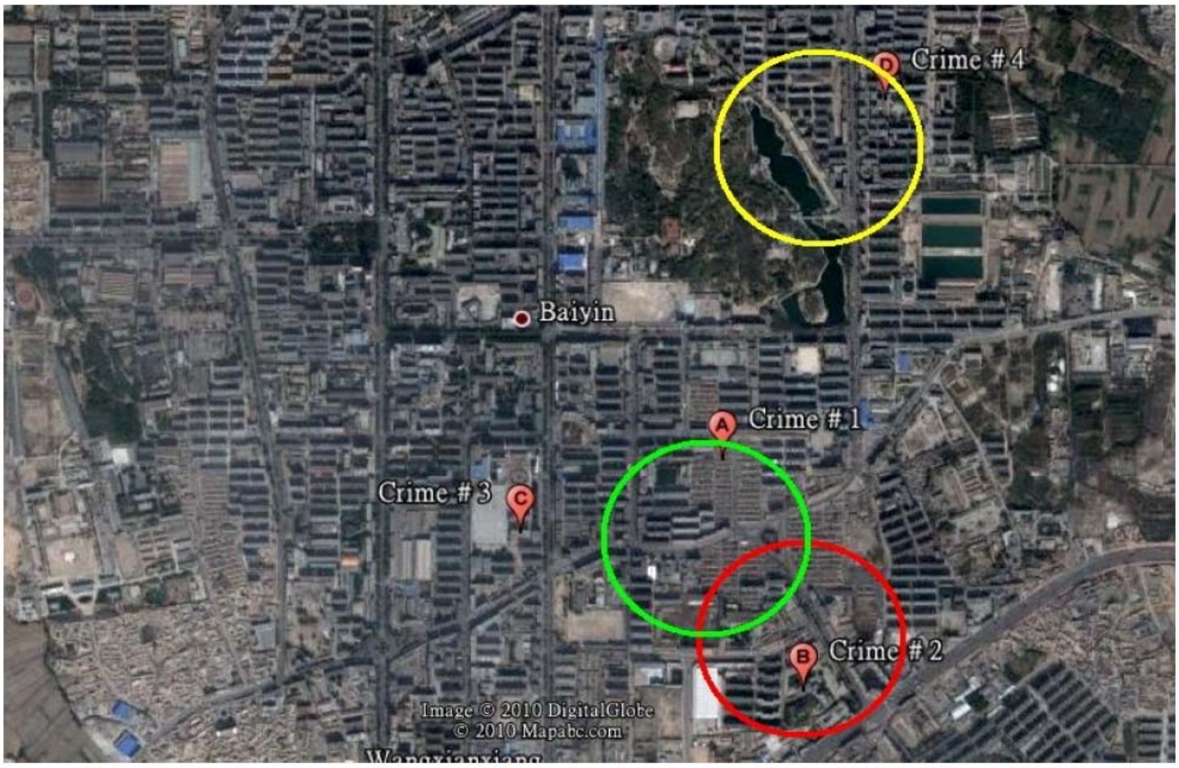

Figure 7. The result of prediction.

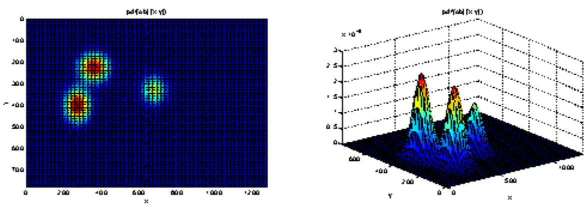

Figure 5. The Geographical Profiling of Hospital Factor.

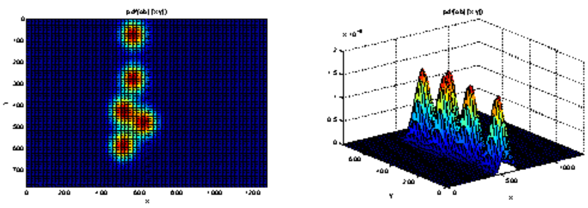

Figure 6. The Geographical Profiling of Hotel Factor.

calculate the distribution in step (3).

(3) Make the geographic profile of each factor selected above by the discrete distance decay function forementioned, like Fig.2-Fig.6.

(4) With the first kth ($k \geq 1$) crime sites known, we combine each factor's geographic profile by the dynamic prediction model mentioned in Part 2.2 and generate the final geographic.

(5) Set the threshold of accept probability of final geographic profile, P_θ. And plot the area where the probability is greater than threshold.

(6) Back to step (4) if k is not the maximum number of crime.

We can see the result of the prediction in the Fig.7. The red circle means the prediction of the second crime with the prior distribution and the first crime site. The green one means the prediction of the third crime site with the first two crime sites. The yellow one represent the crime site prediction of the forth crime. In the figures, we could see the second and fourth crime sites prediction is considerable correct. However, the third crime site is not include in the prediction region.

5 Conclusions and discussion

The result of this model is remarkable and impressive. Though we can not give prediction of time of next crime, the given predicted area will help local police a lot to arrest criminals.

Limited to the lack of crime data, we can not do a great deal of computer simulation which will help us to study how the empirical parameters affect the result. However, to further evaluate our model, we put forward the utility function as a framework of evaluation:

$$F(S,X)=\sum_{i=1}^{n}\frac{1}{S\cdot(H(X,y_i)+\omega)} \tag{14}$$

where S is the area of prediction region, $H(X,y_i)$ is the tth Hausdorff distance between prediction region and real crime site and ω is a manually set constant which assure the denominator nonzero.

Besides, due to that we have only considered the geographical factors which can affect the prediction of crimes, it may lead to some deviation in our work. So, we should take more factors unrelated to the geography into consideration to improve the accuracy of the model. If we combine these two aspects into our model reasonably, it will become more robust and practical.

6 Future Work

In this paper, we introduce Bayesian learning theory into crime prediction and get a satisfactory result. Nevertheless, our model unavoidably depends on the selection of parameters. In later work, we will try to access more crime data for simulation. To study the effect of parameters in detail and improve our prediction, we will evaluate results in the framework of forementioned utility function and try some heuristic algorithms to optimize the parameters. Moreover, we will quantitatively study the factors unrelated to geographical distribution.

Acknowledgements

This work is partially funded by the NCET from the Ministry of Education of China.

References

[1] J. L. LeBeau, "The Methods and Measure of Centrography and the spatial Dynamics of Rape", Journal of Quantitative Criminology, Vol.3, No.2, 1987.

[2] D. Weisburd, W. Bernasco, G. J.N. Bruinsma, "*Determining How Journeys-to-Crime Vary: Measuring Inter- and Intra-Offender Crime Trip Distributions*", Springer New York, 2009.

[3] L. E. Cohen, M. Felson, "Social Change and Crime Rate Trends: A Routine Activity Approach", American Sociological Review, Vol.44, 1979.

[4] D. Canter, P. Larkin, "The environmental range of serial rapists". Journal of Environmental Psychology, 13, 63C69, 1993.

[5] D.K. Rossmo, "*Geographic Profiling*", CRC Press, New York, 2000.

[6] D.K. Rossmo, "Geographic heuristics or shortcuts to failure?: Response to Snook et al.", Appllied Cognitive Psychology, 19, 651C654, 2005

[7] J. Stile, D. Brown, "Geographical profiling with event prediction", IEEE International Conference on Man and Cybernetics, 3712-3719, 2003.

[8] G. C. Oatley, B. W. Ewart, "Crime analysis software: 'pins in maps', clustering and Bayes net prediction", Expert Systems with Applications, 25, 569-588, 2003.

[9] D. Canter, T. Coffey, M. Huntley, C. Missen, "Predicting Serial Killers' Home Base Using a Decision Support System". Journal of Quantitative Criminology, Vol.16, No.4, 2000.

[10] *http://news.xinhuanet.com/comments* $/2004-12/18/$*content_*$2350051.$*htm*