Applying Spatiotemporal and Demographic Data to Locate Next Crime Location

Control #7501

February 22, 2010

Abstract

Geographic profiling is a tool used by law enforcement to predict the location of a serial criminal's next crime. A typical geographic profile outputs estimated probabilities with the input of time and location of previous crimes. In this paper, we develop a new geographic profile that is able to incorporate demographical observations while remaining an effective predictor. We assume that (1) there are buffer zones around previous crime scenes because the criminal fears capture, (2) there is distance decay as criminals prefer something about the locations where previous crimes were committed, and (3) criminals target potential victims based on income and (4) target areas based on crime rate, which are claims supported by research of serial criminals. In order to find an effective profile, we have combined two models of criminal behavior which predict the location of future crime. First, we compute probability densities using a time-weighted kernel density algorithm, which includes buffer zone and distance decay functions. We test this model and find it does well with respect to a control algorithm. Second, we utilize Markov chains to model the criminal's attraction to certain neighborhood characteristics: income and crime frequency. These two profiles are combined into one algorithm by utilizing the time-weighted kernel density algorithm to identify locations at high risk of future crime, then applying the Markov model to further refine our predictions. We then apply our model to the notorious case of the Boston Strangler. The predictions of the combination algorithm are compared against the predictions of a standard control algorithm. Our combination algorithm does not do well in predicting future crimes. We discuss the effects of these weaknesses on the expected effectiveness of the model in predicting future crime locations. Included also is an executive summary for the chief of police.

Contents

List of Figures

List of Tables

1 Introduction

Everyone knows the profile of a serial criminal. He (most are male) is brilliant and moody. He targets those around him, killing them in imaginative and horrific ways. He plays extended mind games with the police, leaving clues in the pattern of his many killings. He looks like Anthony Hopkins.

But everyone is wrong. In fact, the average serial criminal generally is not a mastermind. Often his victims are complete strangers. And he generally does not leave clues in the locations of his killings [1]. Most importantly, there is no such thing as the 'average' serial criminal. We define a serial criminal to be a criminal who steals, rapes, commits arson, assaults, or kills on at least three times.

With the advancement of computational technology, new methods of investigating serial crimes have been developed. Police departments have started to incorporate geographical profiling software to help in their investigations of serial criminals. The geographic profile of a criminal can lead police to the criminal's home base or the next probable location simply based on previous crime locations. Considerations beyond previous crime locations may also be taken into account. Next crime location estimates would allow the police to focus on a relatively small area and hence cut short the time that criminals are free to victimize the innocent.

This technique of geographical profiling has been implemented successfully in cases like that of serial killer Peter Sutcliffe, who attacked and killed at least 13 people by 1981. Since then, more advanced software programs such as CrimeStat, Dragnet, and Rigel have been developed. An effective geographical profiling software program is appealing because it is useful to the police and it has a very limited number of input variables (e.g. time and location of crime).

The current software do not generally take into consideration anthropological observations. A serial criminal usually targets similar victims and often will target them in similar locations[1]. Although it requires more information on demographics than existing software programs, we will attempt to incorporate factors such as income and crime rates of the area. We will combine this with a spatio-temporal algorithm already used in current software to create a new model for geographic profiling.

1.1 Plan of Attack

Our aim is to create an algorithm that effectively aids law enforcement in capturing serial criminals. In this paper, we will

1. Define what we mean by 'effective algorithm'.
2. Present two models of criminal behavior capable of defining an area within which the next crime is predicted to occur.
3. Combine these results to obtain our algorithm.
4. Test our algorithm against the notorious case of the 'Boston Strangler'.

1.2 What Makes a Good Algorithm?

A geographical profiling algorithm produces a probability density function corresponding to the probability that the crime will occur in some area. An effective algorithm has three basic features.

1. It is accurate. In order to be useful, our algorithm should accurately predict the site of future crimes better than a selected control algorithm.
2. It is compact. The area defined by the algorithm should be smaller than a selected control algorithm. If the predicted area is too large, it is unlikely to be useful to law enforcement.
3. It is fast. By fast, we mean that the criminal is caught sooner rather than later. The cost associated with repeated failure is high, and any definition of effective should weight early success more than later success.

With these criteria in mind, it is time to develop a metric we will use to evaluate our algorithm with respect to a control algorithm.

1. The period of time after the discovery of the 3rd crime will be simulated..
2. After the crime occurs, the predicted probability density of our algorithm will be compared to the predicted probability density of the control algorithm. The difference between these numbers will then be weighted by time and summed.
3. We will also examine statistics on whether the crime lay within a predicted boundary, and how wide the predicted boundary was compared to the control boundary.

1.3 Basic Theory of Geographical Profiling

The basis of geographical profiling is environmental criminology, which has several foundational theories [2] which we will use in our models.

Crime pattern theory supposes that crime is not random. The location of a crime is likely near a criminal's normal activity space. The normal activity space is the collection of areas where the individual most frequently comes into contact with others.

Routine activity theory supposes that crime is a function of the target and presence of guardians. In other words, in order for crime to happen, there must be a suitable target and no guardian. The target must be one that makes the offender motivated enough to commit the crime, and the time and location must be such that there are no police or other guardians available to stop the crime. We interpret this to mean the criminal prefers certain types of targets.

Rational choice theory This theory is about the criminal's decision making. It says that criminals make choices that benefit themselves (i.e. they are rational).

2 Control Algorithm

The control algorithm is selected to represent the traditional 'center of mass' method of predicting serial crime [4]. Given a set of points corresponding to the crime locations, we compute the average point and the standard deviation of the points. A two dimensional normal distribution with the computed average and standard deviation is then generated to represent the probability that the crime will occur at a location.

3 Time-Weighted Kernel Density Model

Our first algorithm makes use of a method known as a time-weighted kernel density function, which was shown to be the most effective among a selection of algorithms currently in use [4].

3.1 The Criminal Mind Examined

We will make the following assumptions in line with previous research on criminal behavior [2].

1. The criminal does not wish to return to the scene of the crime. Instead, there exists a buffer zone surrounding each crime scene. The criminal fears capture and will tend to avoid the buffer zone. In other words, we assume that their repulsion from the scene of the crime is proportional to the distance from the crime.

2. However, using crime pattern theory, the criminal will feel some drive to commit the next crime near a previous crime scene. In other words, we assume that their attraction to previous crime scenes is inversely proportional to the distance from the crime (distance decay).

3. As time passes, these first two assumptions are weakened with respect to each crime scene – the criminal has more means of mobility and also is less fearful of capture near the original crime scene.

4. The criminal attempts to maximize his own utility (rational choice theory), which is a function of their distance from the crime scene.

With these assumptions in mind, we now derive our model.

The *area of interest* is a set of points $A = [x_1, x_2] \times [y_1, y_2]$ in $\mathbb{R}^2$. This corresponds to the grid which we will examine. Our *information set* is a finite set Ω of points in $A \times \mathbb{Z} = \{[x_1, y_1, t_1], [x_2, y_2, t_2], \ldots, [x_n, y_n, t_n]\}$, where n is the number of crimes committed and t_i is the time at which the ith crime was committed. We want to define a probability density function $\rho : A \to \mathbb{R}^+$. The density function will be of the form

$$\rho(x, y) = \sum_{(x_i, y_i, t_i) \epsilon \Omega} t_i^w f(r),$$

where f is a probability function of the distance between (x, y) and (x_i, y_i), and t^w represents a weighting function that puts more weight on more recent crimes. Here, $r(x, y) = \sqrt{(x - x_i)^2 + (y - y_i)^2}$. Consider (x, y) to be the position within the area of interest and (x_i, y_i) to be the position of the ith crime, where $1 \leq i \leq n$ and n is the number of crimes in the series.

The function f is determined by our first two assumptions. These two assumptions deal with the criminal's attraction and repulsion to certain locations. Assumption 1 implies that criminals feel a repulsive force when close to a scene of their previous crime. However, assumption 2 implies that criminals also feel an attractive force when close to a scene of their previous crime. Let $U : \mathbb{R}^+ \to \mathbb{R}^+$ be defined by $U(r) = a * r + c/r$, where U is the dissatisfaction of the criminal when distance r away from a previous crime scene, and a and c are constants. Using assumption 4, we predict that the criminal will minimize this function. The function is minimized where $\frac{dU}{dr} = 0$, or where $a - c/r^2 = 0$, which is when $r = \sqrt{c/a}$. Then our model predicts that the criminal has an optimal distance to travel from crime scenes.

In order to generate error bounds, we will use a normal distribution function centered at previous crime scenes, or

$$f(r) = \frac{e^{\frac{-r^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}},$$

where σ is the standard deviation of distance between previous crimes.

3.2 Are These Assumptions Realistic?

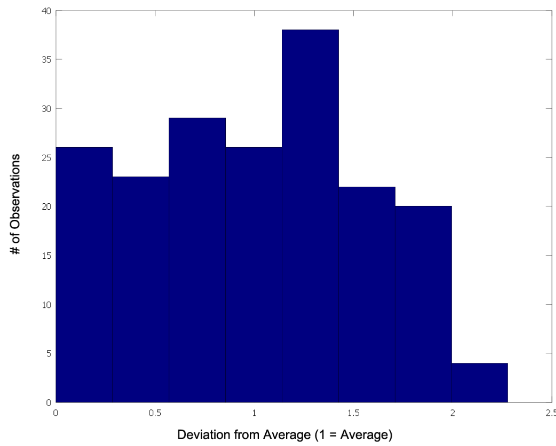
We aggregated our data on serial criminal location. The final data contain serial killer and serial rapist information. The distance between different crimes were normalized and plotted in Figure 1. The data appear to support our assumptions. Closely situated crimes are a less frequent occurrence, while the frequency of crime tends to drop off sharply past a certain point. The data also support assumption 4. Often, a crime will occur later in nearly the same location, but successive crimes rarely occur in the same location.

The data also suggest that the criminal is likely to chose an optimal distance greater than the average distance between crimes, specifically, 1.25 times greater. This suggests that a normal distribution is not optimal, and that a skewed distribution would be more effective. However, we will use a normal distribution due to ease of implementation.

3.3 Time-Weighted Kernel Density (TKD) Algorithm

We simulated the probability density function by numerically computing discreet probability values at each pixel on a map of the area of interest. We used the following method:

Figure 1: Distribution of Distances Between Crimes



1. Each known crime is assigned an integer value to represent time in days since the first crime.
2. For (x, y), compute

$$\rho(x, y) = \sum_{(x_i, y_i, t_i) \epsilon \Omega} t_i^w f(r)$$

3. Repeat for all (x_i, y_i), where $1 \leq i \leq n$.

This algorithm produced the correct output.

3.4 Case Study: Jack the Ripper

To see how the algorithm would have performed in a real world situation, we imagine that we are transported through time and space to London, 1888, when Jack the Ripper terrorized the foggy streets of Whitechapel. We compiled data on the location and time of each killing and evaluated our algorithm. Values in the 'Control' and 'TDK' column represent the predicted probability of the crime occuring in the neighborhood of the area it did occur in.

Our algorithm did well with respect to the control algorithm, partly because the Ripper killings tended to be widely spaced and unpredictable, something which the control algorithm has trouble handling. Early crimes were actually predicted better than later crimes. Unfortunately, as can been seen in Table 2, the total predicted area comprised a significant percentage of the area of interest, around 40 - 50% at 50% confidence.

Figure 2: Blue points represent previous crimes, while red represents the current crime.

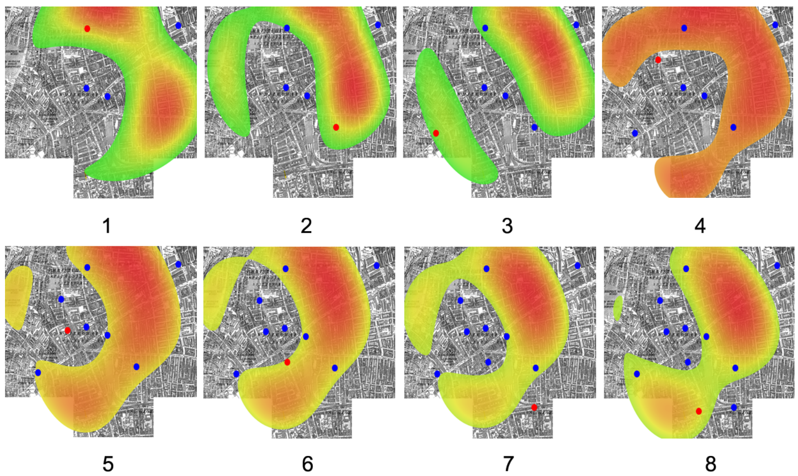
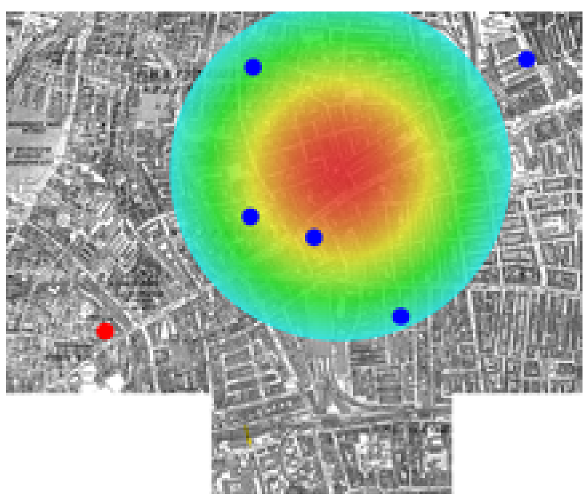


Figure 3: Control Prediction



4 Markov Model

Now we take a different approach. In the first model, we assumed that the serial criminal's behavior was dependent on the location of previous crimes. In contrast, here we assume that the criminal tends to prefer certain areas which

Table 1: Comparison Between TDK Algorithm and Control Algorithm

Crime #	Control	TDK	Difference	In 75%?	In 50%?	In 25%?
4	0.57	0.88	+0.32	yes	yes	yes
5	0.42	0.80	+0.38	yes	yes	no
6	0.14	0.72	+0.58	yes	yes	no
7	0.69	0.89	+0.19	yes	no	no
8	0.83	0.72	-0.11	no	no	no
9	0.81	0.80	-0.01	yes	no	no
10	0.30	0.74	+0.44	yes	no	no
11	0.49	0.85	+0.36	yes	yes	no

Table 2: Percentage of Area of Interest Predicted With 50% Confidence

Crime #	4	5	6	7	8	9	10	11
Percentage:	41%	39%	41%	46%	45%	43%	43%	49%

exhibit certain demographics. The criminal may prefer these areas for personal reasons or because it is easier to commit the crime in these areas. This is an assumption supported by research [1] [8].

4.1 The Criminal Mind Re-Examined

The specific demographics we will consider will be income and crime. Our assumptions are the following.

1. The criminal first chooses a base location for a crime according to factors not reflected in this model.
2. The criminal then is attracted to certain demographics and tends to move away from the original base location towards these areas.

With these assumptions in mind, we derive our model. This is similar to models used in the biological sciences to model predator movement [7].

Our area of interest is now an $n \times n$ real valued matrix A. Our *information set* is a finite set $\Omega = \{(x_1, y_1), (x_2, y_2, \ldots, (x_n, y_n)\}$, where x_i and y_i are integers between 1 and n and n is the number of previous crimes committed. Here we have split the area in which the crimes are being committed into an $n \times n$ grid with information in each square corresponding to demographics such as income or crime.

Now we define the criminal's preference to be a real number representing the demographic to which the criminal will be attracted. We do not know the criminal's preference, so we use the criminal's *observed preference* P, where

$$P = \frac{\sum_{(j,k)\in\Omega} A_{jk}}{N},$$

with A_{jk} the j,kth entry in A. In other words, the criminal's observed preference is the average of the values where the criminal had previous committed the crime. We also compute the standard deviation σ of the observed values where the criminal previous committed the crime. Given the criminal's observed preference, we can define a utility matrix U for the individual by

$$U_{jk} = e^{\frac{A_{jk}-P}{2\sigma^2}}.$$

The utility function measures the difference between the criminal's observed preference and the area of interest at some point (j,k).

We implement our assumptions in the following way. Assume that the criminal is located at some entry s in the area of interest. Then the probability that they will move to a neighboring location n is

$$\frac{U(n)}{\sum_{s'} U(s')},$$

where s' is the set of all neighboring entries. Once they have migrated to that position, the process repeats itself indefinitely. The problem now becomes the following: Given some initial location (x,y), predict the probability that the criminal will be in a certain location after some number of steps.

4.2 Markov Chain Algorithm

In this model, the probability that the criminal will be in a certain location at some future time depends only on the current location of the criminal. So to predict the movement of the criminal we use Markov chain methods. Our area of interest is an $n \times n$ matrix, so we have n^2 possible locations for the criminal's next crime.

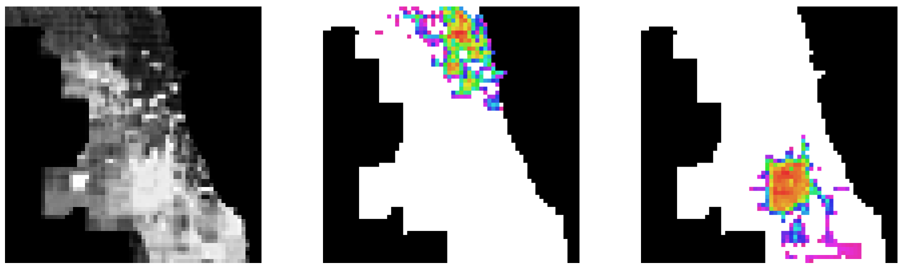
We define an $n^2 \times n^2$ matrix B where B_{ij} is the probability of moving from state i to state j. Each state represents a possible location for the criminal. We already defined the probability of moving from state i to state j if the two locations the states represent border each other. To find the probability that the criminal is in a certain location after m steps we compute B^m and multiply it by x. This will yield the desired result [6].

We implemented these steps in an algorithm using data from the city of Chicago on crime. Chicago was divided into a grid 64 by 64 units, and then an initial location was simulated with several initial preferences. The results are shown in Figure 4. As expected, our algorithm showed that given some initial location the criminal was more likely to move towards their preferred demographic.

5 Combination Algorithm

Next we develop a combination algorithm. The combination algorithm combines the TKD and Markov algorithm. First, the TKD algorithm is run. An average

Figure 4: Markov Model Simulation Results



(a) Original Income Map

(b) Criminal Prefers High Crime

(c) Criminal Prefers Low Crime

Table 3: Comparison Between TDK Algorithm and Control Algorithm

Crime #	Control	Markov Model	Difference	In 75%?	In 50%?	In 25%?
4	0.35	0.32	-0.02	yes	no	no
5	0.78	0.47	-0.31	yes	no	no
6	0.80	0.50	-0.31	yes	no	no
7	0.50	0.62	+0.13	yes	yes	no

probability is computed for each entry in the Markov algorithm's area of interest. The Markov model is then run, using the most recent crime location as the starting location for the algorithm.

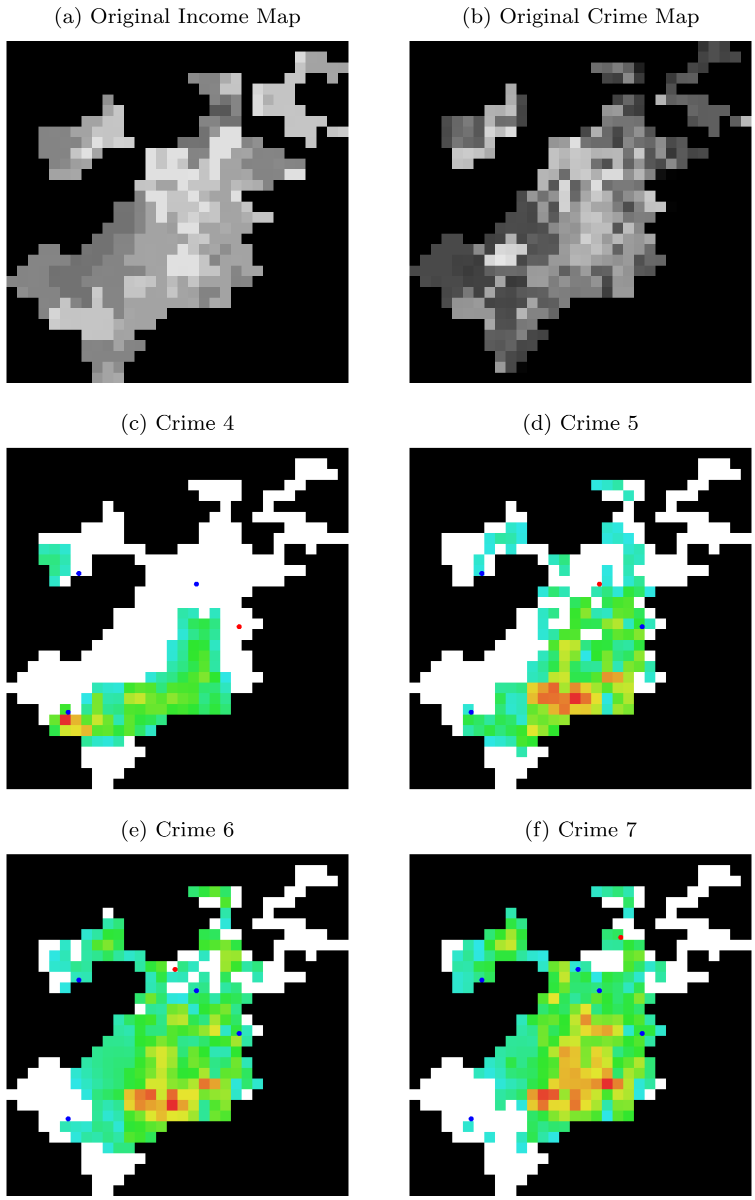
Using data on income and crime in Boston, we tested our algorithm on the case of the Boston Strangler, who stalked the inhabitants of Boston in the early 1960s, committing a series of seven crimes in the Boston area.

5.1 Case Study: The Boston Strangler

Summarized in Table 3 and Figure 5 are the results of the analysis of the Boston Strangler killings.

The combination model performed markedly worse than the control. The control was a better predictor in all but the last crime, although the combination algorithm did contain the next crime location in its prediction with 75% confidence. Clearly we need more data with which to test our model. The poor performance of the combination algorithm suggests that more complex is not neccesarily better. However, one factor which could have affected the performance of the algorithm in the Boston case is the fact that the crimes were not necessarily widely spaced. For example, the 6th crime occured a block away from the 1st crime. So the assumptions of our first model were not entirely satisfied.

Figure 5: Combination Model Simulation Results



(a) Original Income Map

(b) Original Crime Map

(c) Crime 4

(d) Crime 5

(e) Crime 6

(f) Crime 7

6 Conclusions

In general, our model does better when the criminal spaces their crimes and targets certain demographics. However, we need more data in order to more completely test our model. Given our tests so far, we recommend that our model by applied only with great caution.

7 Strengths and Weaknesses

Strengths The model takes into account data other than location of crime.

Weaknesses

1. The assumptions of our model are not necessarily very true. For example, an intelligent and mathematically literate criminal who simply throws darts at a map on the wall to determine the location of their next crime is not modeled at all well by this model. If knowledge of the assumptions of the model were to become widely known, it would invalidate the model.
2. If the criminal does not have a preference for certain demographics the Markov model becomes irrelevant. However, by construction, the overall model simply reverts to a rough approximation of the original TKD Model.
3. The further the distance between crimes, the more generally difficult and inaccurate the model becomes. In the case of a single city, the numbers are manageable. In the case of, for example, a large state county, the time required to compute increases exponentially. And the total spread of the predicted area increases as well.

Avenues for Improvement

1. As noted in the section on the TKD Model, empirical data suggest that the distribution of distances between crimes is a skewed distribution, not a normal distribution. It may be more accurate to include a different distribution, or even to tailor the distribution on a case-by-case basis.
2. Other types of data could be relevant to the location of the serial killer, for example police coverage or ethnic makeup. The model can easily be expanded to include these factors.

8 Executive Summary

Geographical profiling is not a recent invention. Ever since the first police detective pushed his first pin into a map to mark the location of a murder, police departments around the world have been using geographical profiling to catch criminals. Now computers have improved and promise to improve further this process.

How successful have they been so far? The answer is mixed. On the one hand, the simplest computer model's prediction often is surprisingly accurate when compared to human prediction. However, computer predictions are not yet a substitute for old fashioned police work; at best, they provide a means of deciding where to allocate scarce departmental resources. The problem is that there is no such creature as an 'average' serial killer, and behavioral predictions are therefore very imprecise.

However, the future holds some promise for improving these models, not necessarily by changing the models themselves but by incorporating new data. The increasing proliferation of spatial statistics – data on income, crime, ethnic makeup – could allow for more accurate targeting of future crime scenes.

There are three aspects of geographical profiling which need to be considered in order to create a good algorithm. First, and most obviously, it should be more accurate in predicting future crimes than previous methods.

Second, it should not produce an overly broad area. If the area in which the crime is predicted to occur is too large, it will be of little use to law enforcement.

Third, it should be fast. The sooner the serial criminal is captured, the less damage occurs, and therefore a more successful algorithm will catch the criminal early. Since apprehending a criminal early means doing so with limited data, a good model should be able to work effectively even with minimal data.

With these issues in mind, our team built a model which incorporated both location of previous crimes and demographic characteristics of the area where the crime occurred. Our model was based on two models: one based on the location of previous crimes and one based on demographics.

The first model made two important assumptions about the behavior of the criminal, which are often used in geographical profiling. First, we assumed that the criminal would be reluctant to return to the scene of the crime (at least until some time had passed) out of fear of capture. Second, we assumed that there was something attractive to the criminal about the scene of the crime and that if capture was not a concern, they would return to the same location the commit the crime. Using these two forces as a guideline is enough to generate a prediction about the next crime.

But not necessarily a useful prediction. Although in our simulation the first model did well in predicting the location of future crimes – it accurately predicted the location of the crimes with 75% confidence – the total area with high probability of crime tended to be large, making it a less feasible solution to the problem.

As a result, a second model, the Markov Model, was created. This model ignored the method through which the criminal chooses the initial location of their crime. However, this model made an important assumption – that once the criminal choses the location of the crime they might find it unsatisfactory. Perhaps the criminal prefers certain types of areas, and hence will tend to migrate towards those areas. This is a model used mostly in the biological sciences to model the movements of predators who search out richer areas filled with easy prey.

While it may seem a bit unusual to compare a lone human serial criminal to the movements of packs of wolves, the idea makes some sense. Psychological profiles show that serial criminals are often attracted to certain areas. And attraction is not the only force acting on the criminal mind. Fear of capture may drive the criminal away from areas which are more heavily policed, although our model did not incorporate this aspect of the criminal mind due to lack of data.

The two models were then combined. Although data were difficult to obtain, we tested our final, combined model against the historical case of the Boston Strangler. Unfortunately, it performed poorly when compared to the control algorithm. This could be due to a number of reasons. For one, the crime pattern of the Boston Strangler was bolder than we assumed the criminal would be comfortable with – his sixth murder occurred only a few blocks away from the first one. Second, our predicted area was smaller than the control algorithm. We would need to test our model against more data and in a wider variety of situations in order to attain a more complete grasp of the issues involved.

In general, our model does not perform well when there is not adequate data available for factors such as crime and income, or when the criminal does not prefer certain demographics. Our model also does not perform well the more randomly the criminal acts.

At most, the model should be used as a general guideline for uncovering issues which may not have been considered. It also could be used as a systematic way to reduce the total area under consideration for the next crimes. However, be wary of using it to predict the location of the next crime – earlier and simpler models do a better job of that.

9 Data Sources

Unfortunately, data on this subject is hard to find. Police departments are reluctance to release information on serial criminals without a background check due to privacy concerns and public relation concerns [3]. We compiled information on the data and location of the crimes of the Boston Strangler and Jack the Ripper, and pulled additional information on serial criminal crimes and location from http://www.icpsr.umich.edu/icpsrweb/CRIMESTAT/.

References

[1] Egger, Steven. The Need to Kill: Inside the World of the Serial Killer. Prentice Hall, Upper Saddle River, NJ. 2003.

[2] Tania, Pal. Accuracy Assessment of Individually Calibrated Journey-to Crime Geographic Profiling Models. Master's Thesis. 2007.

[3] Paulsen, Derek. Crime Mapping and Spatial Aspects of Crime. Pearson Education, Upper Saddle River, NJ.

[4] Paulsen, Derek. Predicting Next Event Locations in a Crime Series using Advanced Spatial Prediction Methods. 2005. Third National Map Conference.

[5] Nickerson, R.S. Bertrand's chord, Buffon's needle, and the concept of randomness. Erlbaum Taylor and Francis, London, UK.

[6] Pardoux, Etienne. Markov Processes and Applications. John Wiley and Sons, West Sussex, UK.

[7] Johnson, Shane; Bowers. Stability of Space-Time Clusters of Burglary, British Journal of Criminology Volume:44 Issue:1 January 2004

[8] Wood, Henry. Arson. A Geographical, Demographic and Motivational Perspective. Master's Thesis. 1999

[9] Deming, W. Edwards. Statistical Adjustment of Data. Dover Publications Inc., New York, 1964.