A GENERATIVE MODEL OF OFFENDERS’ SPATIAL BEHAVIOUR

RICHARD BACHE
Department of Computer and Information Sciences,
University of Strathclyde, Glasgow G1 1XH, Scotland, UK
richard.bache@cis.strath.ac.uk

FABIO CRESTANI
Faculty of Informatics, University of Lugano, Lugano 6900, Switzerland
fabio.crestani@unisi.ch

Received 5 February 2003
Revised 19 July 2011

The relationship between distance travelled to an offence and frequency of offending has traditionally been expressed as a (downward-sloping) decay function and such a curve is typically used to fit empirical data. It is proposed here that a decay function should be viewed as a probability density function. It is then possible to construct generative models to assign probabilities to suspects from a set of known offenders whose past crimes are stored in a police data archive. Probabilities can then be used to prioritise suspects in an investigation and calculate the probability of being the culprit. Two functional forms of the decay function are considered: negative exponential and power. These are shown empirically to outperform a basic model which simply ranks suspects by distance from the crime. The model is then extended to include also preferred direction of travel which varies between offenders. If direction of travel is incorporated then predictions become more accurate. The generative decay model has two advantages over a basic model. Firstly it can incorporate other information such as past frequency of offending. Secondly, it provides an estimate of suspect likelihood indicating the trustworthiness of any inference by the model.

Keywords: Criminology; offender behaviour modelling; spatial data; generative model.

1. Introduction

There have been many studies that have sought to elucidate spatial patterns within criminal behaviour. However, one rapidly expanding subset of this work focuses not on aggregate spatial distributions but on the spatial movement and location choices of the individual offender and how this can be harnessed in the police investigative context to identify perpetrators of unsolved offences. Much of this work is qualitative in nature but there have been several attempts (cited below) to identify quantitative relationships. This is predicated on the assumption that
there are common underlying patterns in the behaviour of all offenders. Three particular themes emerge from this work:

(i) Identifying the decay function which describes the likelihood of offending as distance from the base of the offender increases.\(^7,12\)

(ii) Linking an unsolved crime to solved crimes or offenders’ home bases by means of distance.\(^4,11\)

(iii) Applying distance decay functions to identify the area to search for the home base of a single unknown offender believed to have carried out a number of similar offences with known locations.\(^9,10\)

The location of crimes and offenders’ home bases is essentially a two-dimensional problem, although the use of a decay function seeks to reduce this to a single dimension. This paper proposes generative models based on decay functions (Generative Decay Models) as a basis for assigning a likelihood measure for suspects for an unsolved crime and thus brings together themes 1 and 2 thus informing the development of approaches pertinent to theme 3. A likelihood measure can then be used to prioritise suspects as part of any investigation. In doing so we introduce a probabilistic approach to an area which hitherto has been concerned with simple curve fitting. However such a model is direction free in that it only takes into account the distance travelled and implicitly assumes that an offender is equally likely to venture from the home base any direction. A refinement of the decay model incorporates preferred direction of travel and thus considers the problem of modelling offenders’ behaviour as truly two-dimensional. These models have the advantage of being sufficiently flexible to incorporate other information relevant to suspect prioritisation such as past frequency of a suspect’s offending.

The work presented here was part of the UK Government funded project, iMOV (Interactive Modus Operandi Visualisation).

1.1. Statement of the problem

We assume an existing archive of solved crimes of a particular type where the location of the offence is known and offender’s base is either known or can be estimated. For some unsolved crime of the same type, we wish to allocate probabilities to known offenders, drawn from the set of offenders held within a police database in a given police jurisdiction and during a given time period, as suspects for the unsolved crime. Furthermore, we may have additional information about the suspects such as the propensity to offend based on, for example, the past frequency of offending. If, for example, an unsolved crime is equidistant from two suspects’ bases and we ignore any directional effect the probability of the most prolific offender would be higher. We want to incorporate additional information into the model in a theoretically principled way.

Previous studies of decay function have used such functions descriptively as a way of exploring the relationship between distance travelled and frequency of
occurrence of offences across a sample. They also ignore any directional effect. Here we wish to integrate the decay function into a predictive model and to do this we need to be able to apply it in reverse to the way it is normally used - given an unsolved crime, how likely is it that an offender operating from a base would travel this far? Such a model may then be extended by taking into account an offender’s preference to travel in a particular direction.

As we shall see below, merely finding a curve which fits the data is not sufficient; there are certain properties a decay function must have if it is to remain stable when applied to a generally useable model. For this reason, we embark on creating a formal definition of a (direction-free) decay function and identify the necessary mathematical properties such a function ought to have. It is then possible to create a directional decay function where the probability of offending depends both on the distance and direction travelled.

1.2. A basic distance-based model

It is worth noting here that we do not actually need a decay function to prioritise offenders. We consider as an alternative the following basic model. If we know the base of each offender (which may be the place of residence or some other location from which he travels to offend) and also the location of the unsolved crime we can simply rank the suspects according to the distance from the offence where the nearest is the most likely to offend. Such an approach has three limitations:

- It assumes a simple (inverse) rank correlation between frequency of offending and distance travelled. This means it makes no allowances for variations in the rate of decay within distance intervals, nor, particularly, for increases in likelihood which have been hypothesised both very close to home (in the guise of a “buffer zone”) and at the longer distances (as a subset of heightened activity focused on specific targets). Moreover, it means the basic model produces rankings but gives no indication of suspect likelihood.
- There is no obvious way to incorporate distance travelled into such a model.
- There is also no way to use additional information. For example, if offender A is marginally closer to the crime scene than offender B, the model will rank A over B even if B is a far more prolific offender than A.

We will therefore use this basic model as a baseline and demonstrate that the Generative Decay Models perform consistently better and addresses all three of these limitations.

1.3. Structure of paper

The remainder of the paper is set out as follows. Section 2 provides a formal definition of a decay function and determines what properties such a function should have. From these properties we can rule out certain functional forms previously
used as decay functions. In Sec. 3 we construct a generative model using those decay functions which possess the necessary properties. Such a model estimates the probability that an offender with a given base will commit a crime in a particular location and then use this to determine the probability that an unsolved crime was committed by one of a number of offenders. Section 4 seeks to incorporate direction into the model by modifying the decay function. Firstly we show that offenders tend to travel in a particular direction by analysing actual data. Secondly we show how an estimate of preferred direction can be incorporated into the generative model. In Sec. 5 we provide some empirical evidence that the Generative Decay Models do predict offenders of unsolved crimes more effectively than the simple model and furthermore it is possible to estimate the optimal parameters for each model. We provide evidence that there is a directional effect and then show that incorporating direction into the model improves performance. Section 6 provides conclusions.

2. Decay Functions

These functions have traditionally been attempts to fit a curve to empirical data. For a set of solved crimes, usually in a particular locality, the distance between the offender’s base and the location of the offence is measured. Here, we can consider a number of possible distance measures such as Euclidean distance, distance travelled by foot or even time to travel; all are consistent with the notion of a decay function. Data points are allocated into distance ranges (e.g. in bins of 100m width) and distance is plotted against frequency. The decay function is an attempt to fit frequencies with distance, with the usual assumption that frequency will diminish with distance. Various authors have attempted to fit different curves. Canter and Hammond\textsuperscript{7} attempted to fit linear, quadratic, log and exponential curves. Turner\textsuperscript{12} used a power curve. However, the purpose of fitting such a curve is to characterise the underlying relationship between the two quantities across a sample rather than to make inferences about unsolved crimes. It is perhaps for this reason that such empirical approaches do not address the required mathematical properties such a curve should have. These problems have to be addressed if we wish to use the function predicatively. In this section we will provide a formal definition of a decay function and consider its properties and their implications.

2.1. \textit{Definition and properties of a decay function}

The fundamental idea behind a decay function is that offending becomes less likely as distance from the offender’s base increases. A generative model assigns probabilities to the various locations for each suspect and then uses Bayes’ Theorem to invert the probabilities so that for a given observation, a probability is assigned to each suspect. We therefore define a decay function $\phi$ as a \textit{probability density function} (pdf) which is a function of distance $s$. As a pdf there are two mathematical properties this function must satisfy: firstly it must always be non-negative:

$$\phi(s) \geq 0$$ (1)
Secondly, the area under the curve must add to one. For the area to be at the very least finite we must make one of two assumptions: firstly we assume there is a maximum distance \( m \) that can be travelled. Such an assumption is not unreasonable since data is usually collected within a bounded area such as a police district. There is in any case a maximum distance an offender can travel without leaving the earth, although setting \( m \) to be half the earth’s circumference is not a reasonable assumption.

\[
\int_{0}^{m} \phi(s) ds = 1 \tag{2}
\]

However deciding what value to assign to \( m \) is problematic. An alternative assumption is to consider the limiting case where \( m \rightarrow \infty \). This could be interpreted as where an offender commits most crimes in a small neighbourhood in the middle of a very large city. In any case, we require that the area under the curve be finite and thus address both of these possibilities below.

Brantingham and Brantingham\(^5\) argue that the offender will be less likely to commit offences very close to the base (offenders do not commit crimes on their doorstep), but the most likely distance will be a short distance from the base. Because data points are put into bins, many empirically derived functions do not pick up this effect and thus assume that the mode is at zero distance from the base (or at least is within the first bin). Our general framework here will allow for either possibility by assuming there is a point \( z \) which is the most likely point to commit an offence. So,

\[
\forall s > z, \frac{d\phi}{ds} \leq 0 \tag{3}
\]

Meaning the curve will slope downwards or be flat after \( z \). For the decay functions we discuss below, we assume the special case \( z = 0 \). This is because the granularity of the distance measurements (see below) is such that even if the doorstep effect were to exist, it is unlikely that the data available would pick this up.

### 2.2. Choice of decay function

Various types of function have been proposed in the literature.\(^7\) Although there are potentially an infinite number of possible decay functions which could be proposed, we focus on those which have been used previously.

- Linear \( as + b \) where \( a < 0 \) and \( b > 0 \),
- Quadratic \( as^2 + bs + c \) where \( a > 0 \), \( b < 0 \) and \( c > 0 \),
- Negative exponential \( ae^{bs} \) where \( a > 0 \) and \( b < 0 \),
- Logarithmic \( a \ln(s) + b \) where \( a < 0 \) and \( b > 0 \),
- Power function \( as^b \) where \( a > 0 \) and \( b < 0 \).

Examples of these curves are given in Fig. 1. The appropriateness of these different functions has direct implications for theories of criminal spatial behaviour and the
particular psychological and behavioural processes which determine the distance decay effect. Although there appear to be two parameters for each model (three for the quadratic), since the area under the curve must equal 1, there is only the freedom to choose one parameter (or two in the case of the quadratic function). Of course, any of these functions can be used to fit a particular data set, but problems arise if we consider the necessary properties discussed above.

The linear and logarithmic functions can cut the x-axis, violating condition (1). Indeed in Canter and Hammond’s study\(^7\) the best fitting linear and logarithmic curve cut the x-axis within the range of the data used to estimate the curve. If we interpret the function as a pdf, such a situation causes the model to break down. It is possible to choose values of \(a\) and \(b\) where this is guaranteed not to happen, by forcing the intersect to be at a value greater than the largest distance. But this is crucially dependent on knowing the maximum distance in advance. Furthermore, from Eq. (2), we can see that altering \(m\) will alter the values of the coefficients. It is easy to show that the shape of the curve will be very sensitive to the choice of \(m\).

A quadratic function is problematic for two reasons. Firstly, it may cut the x-axis and this leads to the same problem as the linear and logarithmic functions although this will not always be the case. However, inevitably it will reach a minimum value and start to increase. If this occurs before reaching \(m\), this violates Condition 3. In Canter and Hammond’s study it both cuts that access and starts to increase at distances less than the farthest data point. Thus the problem with this quadratic function is potentially worse than that of the linear and logarithmic ones.

Both the negative exponential function and the power function will always satisfy constraints 1 and 3. Furthermore since the values converge to zero, the effect of changing \(m\) will not greatly affect the total area under the curve. A problem with the power function occurs at a distance zero, since a singularity occurs. Turner\(^12\)
obviates this problem by adding a small positive quantity to all the distances. We propose to do the same. Of course, both functions are strictly decreasing and this implies $z = 0$.

### 2.3. Assuming infinite space

We consider now what happens as $m$ tends towards infinity. Firstly, for the quadratic, linear, and logarithmic it is impossible find any parameters which give a finite area under the curve. The area under the negative exponential curve will always be finite and will be 1 if we set $a = -b$. The area may be finite for the power function if we assume a small positive quantity $\epsilon$ added to all distances.

$$AUC = \int_{\epsilon}^{\infty} \phi(s)ds$$

$$= \int_{\epsilon}^{\infty} as^b ds$$

$$= \left[ \frac{a}{b+1} s^{b+1} \right]_{\epsilon}^{\infty}$$

For this quantity to be finite we would require that $b < -1$. However, when we attempt to estimate $b$ below, we find that values $-1 < b < 0$ appear to work better. It is quite easy to show that if we calculated the mean distance travelled, to ensure that this was finite we would require $b < -2$.

### 2.4. Discussion

We wish to construct a model which can be applied to unsolved crimes and therefore want the parameters which define the decay functions to be fixed in advance. For this reason, the linear, quadratic and exponential functions are likely to give rise to unstable models which are likely to lead to absurdities. Of the five ranking functions considered, only two are really usable in the context of the model proposed here, namely the exponential and power functions. We do not have to assume a maximum distance $m$ and thus this quantity does not appear as a parameter in the model. We note here that we have chosen two simple functions which meet the criteria we have established above. There are possibly many other functions that might be considered such as splined polynomials and the model developed in the rest of the paper is potentially extensible to other decay functions.

### 3. A Generative Model for Suspect Prioritisation without Direction

Generative models have been used in suspect prioritisation when comparing and seeking to match textual descriptions of solved and unsolved offences and offender profiling;\textsuperscript{1-3} the models in question were probabilistic language models. The approach used here is similar in that we estimate the probability that the actions of
a particular suspect could result in an observation and then use Bayes’ theorem to invert the probabilities. The application Bayes Theorem will require an appropriate prior. Thus, in order to estimate the probability that an offender has committed an unsolved crime, we first estimate the probability that a given offender would have committed a crime in such a location.

3.1. Probability of a given location

For each possible distance $s$, there will be a circle of points with the suspect’s base at its centre. Using the fact that the circumference is proportional to the radius we can write down the probability density function for a crime being committed at a location $l$ given that the suspect, $u$, has a base, where the distance $s = s(l, u)$.

$$\text{pdf}(l \mid u) = k_1 \cdot \frac{\phi(s)}{s}$$

We could seek to eliminate the constant $k_1$ by considering the area under the curve must add to 1, but, as we see, the application of Bayes’ theorem means that this is unnecessary.

3.2. Probability of a suspect

We assume that we have a finite set of suspects $U$, as must be the case within any police department working with those offenders held within its databases. We will also assume that initially we are seeking a single culprit. If there were two or more perpetrators, they are likely to travel together from a single meeting point which we assume to be one of the offenders’ bases. Also, once the police have found one culprit it becomes easier to identify accomplices.

Since we are using Bayes’ Theorem approach, we need a prior probability - that is an initial belief in the likelihood of each of the possible suspects, ignoring any geographical information. The simplest prior here is to assume that all offenders are equally likely to have done it and thus assign an equal probability to each; this we shall term the uniform prior. The uniform prior implies that distance information is the only factor used in prioritising suspects. However, one advantage of this Bayesian approach is that we can build into the model other information. If, for example, two offenders have bases equidistant from the location of the crime, we might favour the offender who had the most prolific history of offending. We could thus assign a probability proportional to the number of similar offences; this we term the Casablanca prior (“Round up the usual suspects.”). The prior could be constructed subjectively by a police officer allocating probabilities based on intelligence. It might include, in the instance of an unsolved burglary, an offender who had committed other property crimes and is believed to have advanced to burglary; although this probability would perhaps be less than someone with a history of burglary. We could also build into the prior the period of time since the last known offence to capture the notion that some offenders might have ceased offending, or,
more heuristically, the offender’s age, given the known markedly skewed age distribution of offending across age. These more complex priors are not explored further here except to note that they can provide alternative ways of introducing additional information into the model and are consistent with the model framework proposed here. We shall focus on the uniform and Casablanca priors.

For each suspect, \( u \in U \), we have a prior probability \( p(u) \). If we know the location of the crime and the suspect’s base then we can calculate the probability that the suspect committed a crime at a position \( p \). By using Bayes’ theorem we have:

\[
p(u \mid l) \propto p(u) \cdot \text{pdf}(l \mid u)
\]

\[
\Rightarrow p(u \mid l) = k_2 \cdot p(u) \cdot \text{pdf}(l \mid u)
\]

The measure \( p(u \mid l) \) can be used both to rank suspects in order of priority, but also, we maintain, to indicate how likely each suspect is of committing the offence.

### 3.3. Estimating the model parameters

We can now consider the two types of decay function discussed above. Substituting in Eq. (7) into Eq. (9) and combining the constants, we have:

\[
p(u \mid l) = k_3 \cdot p(u) \cdot \frac{\phi(s)}{s}
\]

For the negative exponential function we have:

\[
p(u \mid l) = k_4 \cdot p(u) \cdot \frac{e^{bs}}{s}
\]

Now, given that a crime has been committed, someone must have done it. Assuming all suspects are known offenders, the probabilities added together will be unity. So,

\[
\sum_{u \in U} p(u \mid l) = 1
\]

This implies

\[
k_4 = \frac{1}{\sum_{u \in U} p(u) \cdot \frac{e^{bs}}{s(l,u)}}
\]

and so there is only one parameter to estimate for the negative exponential model, namely \( b \). For the power function we have:

\[
p(u \mid l) = k_5 \cdot p(u) \cdot s^{b-1}
\]

\[
k_5 = \frac{1}{\sum_{u \in U} p(u) \cdot s(l,u)^{b-1}}
\]

Again, we only have one parameter to estimate for the power model. This can be done heuristically by considering the effectiveness of the model on actual data and varying \( b \) to give optimal performance, which we show below.
The single parameter for the power model is actually dimensionless, which means that it does not depend on the units used to measure distance. This is not true for the negative exponential model.

4. Adding Preferred Direction to the Model

Although it is widely accepted that there is a relationship between distance travelled and probability of offending, the idea of a preferential direction is less well established and certainly not quantified. Here it is helpful to think of a coordinate system where the offender’s home base is the origin. We can express the position of any crime by polar coordinates where each point is described by a distance and azimuth. By convention, an azimuth of $0^\circ$ points East and the value increases counter-clockwise; thus North is $90^\circ$, West is $180^\circ$ etc. We assume that there is some preferred compass direction of travel $\theta$ and thus the probability of the offender travelling in direction $\psi$, $pdf(\psi | \theta)$ will have a maximum value when $\psi = \theta$ and a minimum value when $\psi = (\theta + 180^\circ) \pmod{360^\circ}$ i.e. the opposite direction.

We are not maintaining that all offenders prefer the same direction. Our hypothesis is that each offender has his own particular preferred direction of travel. Suppose one offender’s preferred direction was south then a location 1km south would be a more likely crime scene than 1km north. The probabilities of locations 1km east or west would lie between the two. Attempting to model the probability density over two-dimensional space using a bivariate normal distribution would be inappropriate since this could not capture the concept of preferred direction. In order to assume that the home base is the most likely place a crime would be committed and be consistent with a decay function with $z = 0$, we would also need to assume that an offender would be as likely to go north as south. This is not what we require.

The hypothesis that there is a preferred direction of travel can be tested by assuming a null hypothesis that the direction is entirely random.

4.1. A test for directional preference

By using polar coordinates we can separate the direction entirely from the distance travelled. However we must be aware that the azimuth measure cannot be treated as any usual scalar quantity. For example, suppose we took the mean of the azimuths of crimes for one offender. A mean of $0^\circ$ would mean that all journeys were made due East; a value of $180^\circ$ would be consistent with the offender only travelling due West or in any direction with equal probability.

The centroid of all the crime locations will take into account not only the direction but also the distance so the distance between the centroid and the home base is also not a good measure. Thus we propose to take the centroid of unit journeys. For each crime we take the azimuth but ignore the distance and assume it to be 1 unit of distance. We then take the centroid of the destinations of the unit journeys. Such a measure, by definition, takes no account of the actual distance travelled.
Figure 2 provides an illustration of this. If there is no directional effect, we expect the unit distance centroid to coincide with the home base. Of course it is unlikely to coincide exactly. However if there were a directional effect we would expect its distance from the home base to be greater.

By taking a number of solved crimes where the home base of the offender is known, we can compare the distance of the actual unit distance centroid from the home base to a simulated unit distance centroid where for each crime we assign a random azimuth which is uniformly distributed on the interval $[0, 360)$. By applying the ranked sign test to the difference between the actual distance and a simulated distance it is possible to reject the null hypothesis if the test shows significance. In this case the azimuth of the (actual) unit distance centroid is the estimate of the preferred direction of this offender.

4.2. Creating a directional decay function

We shall assume that the direction chosen by the offender is independent of the distance travelled. We can therefore express the probability that the offence will occur in location $l$ given that suspect was $u$ by considering a distance $s$ and azimuth $\psi$ (using $u$ as the origin) as

$$pdf(l \mid u) = pdf(s) \cdot pdf(\psi)$$

(16)

The first factor is simply proportional to the decay function. For the second factor we need to consider the azimuth and also that as the distance increases the circumference of the circle increases. If all azimuths were equally likely then

$$pdf(\psi) \propto \frac{1}{s}$$

(17)

and indeed this gives us Eq. (7). However if we wish to give a greater weight to values where $\psi$ is closer to the preferred direction $\theta$ then we need some function of
We propose a cosine function which gives 1 where $\theta - \psi = 0^\circ$ and -1 where $\theta - \psi = \pm 180^\circ$. To create a term which provides a non-negative weighting such that the highest weighting is at $0^\circ$ and the lowest at $\pm 180^\circ$ we add 1; $w$ is used as a parameter of the strength of the directional effect where $0 \leq w \leq 1$.

$$pdf(\psi) \propto \frac{1 + w \cos(\theta - \psi)}{s}$$

Putting these together gives us

$$pdf(l|u) = k_6 \cdot \phi(s)(1 + w \cos(\theta - \psi))$$

The constant $k_6$ is again eliminated by the application of Bayes’ theorem.

5. Empirical Analysis of Models

We now apply the models discussed above to a number of crime sets taken from a police data archive. We note here that it is particularly difficult for researchers to obtain such data due to confidentiality of the data. Thus even the small data sets used here represent a rare access to such data. For the direction-free model we show that it performs better than the basic model where offenders are purely ranked by distance from the crime and often has very high predictive power. We are able to compare the two decay functions and also determine the optimal parameter for each. We then seek to determine if there is a directional effect using the test described in Subsec. 4.1. When adding preferred direction to the model we show that this improves performance.

5.1. Description of the data

We considered crimes committed from an inner city district recorded in a police digital archive. These were categorised according to crime type: theft from vehicles, burglary (two sets), criminal damage, damage to vehicles and robbery (excluding armed robbery). All except burglary B were drawn from crimes committed over a four-year period. Home base coordinates could not be obtained but were estimated using the centroid of all known crimes as has been done on a previous study.\textsuperscript{11} The remaining dataset (burglary B) data considered prolific serial burglars over a 14-year period and home coordinates were present thus there was therefore no need to estimate them. Since the home base is essential to determine the preferred direction the directional model could only be validated on one dataset. Only crimes committed by a lone serial offender (i.e. one who had committed at least two crimes by himself) were considered. The number of data points for each set is summarised in Table 1.

For each crime the location was recorded as an Ordinance Survey grid reference. For Burglary B, the crime scene and home base were represented using 10-figure references, specifying the location to the nearest meter. For all other sets, the location of the crime was represented using 6-figure coordinates, placing in within a 100
Table 1. Summary statistics from crime sets.

<table>
<thead>
<tr>
<th>Crime Type</th>
<th>No. Serial Lone-Crimes</th>
<th>Total No. Offender Crimes</th>
<th>No. Offenders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theft from Vehicles</td>
<td>312</td>
<td>121</td>
<td>33</td>
</tr>
<tr>
<td>Criminal Damage</td>
<td>846</td>
<td>182</td>
<td>57</td>
</tr>
<tr>
<td>Damage to Vehicles</td>
<td>252</td>
<td>37</td>
<td>17</td>
</tr>
<tr>
<td>Burglary A</td>
<td>1289</td>
<td>742</td>
<td>174</td>
</tr>
<tr>
<td>Robbery</td>
<td>263</td>
<td>86</td>
<td>33</td>
</tr>
<tr>
<td>Burglary B</td>
<td>1376</td>
<td>1376</td>
<td>83</td>
</tr>
</tbody>
</table>

meter square. Offenders’ home bases were not known and so were thus estimated as the centroid of all other known crimes.

The distance measure used between the crime and estimated base was Euclidean distance, consistent with the investigative psychological argument that travel behaviour is governed by environmental perceptual processes rather than route calculations. However given the granularity of the data, this would frequently mean a distance of zero if both points were in the same grid square. Given that distances are used in division, this would create frequent singularities as well as consistently under estimate small distances. Thus a small offset was added to each distance to ensure that it was non-zero. Experimentally, it was shown that a distance of 100 meters made the models work optimally.

5.2. Measuring effectiveness

Although there are many ways to determine if a model can rank objects effectively, determining whether the model captures the idea of likelihood (how confident we are that this is the offender) is more problematic, yet these likelihoods are relevant to police investigations and the allocation of scarce investigative resources. Thus we now apply a technique which had been shown to work when prioritising suspects based on text descriptions of the crimes. This technique was itself taken from Information Retrieval and adapted. It is based on the concepts of precision and recall.

In Information Retrieval, it is assumed that there is a large collection of documents of which a small number are relevant to a given user’s information need. This need is expressed as a query and some ranking function compares each document to the query to give some score of relevance. In reality, the document can be assessed by hand to determine whether or not it was relevant and very often such a judgement is deemed binary. Thus for a given number of documents retrieved (e.g. the top ten scoring ones), precision is the proportion of documents retrieved which are relevant and recall is the proportion of all the relevant documents in
The collection which were retrieved. Clearly, as the number of documents retrieved increases, recall will rise and precision tends to fall.

The problem of suspect prioritisation is similar to that of retrieval. There is an unsolved crime and a number of suspects and we wish to rank the suspects according to the likelihood each one could be the culprit. For the score associated with each offender to be considered a measure of suspect likelihood, we have to show that over a number of crimes, the actual value assigned (here a probability of the suspect being the sole offender) is a good predictor of who the culprit is as well as ranking the suspects for one crime. For this reason the measures of entire precision and entire recall were used as defined in.\(^1\) Here, each pair of suspect and crime are given a score, in this case a probability, and all pairs are ranked by that probability. For the top \(N\) pairs, entire precision is the number of pairs where the suspect is the culprit divided by \(N\). Entire recall is the number of such pairs divided by the total number of crimes. Note that we have selected crimes with one offender. An aggregate measure, mean average precision (MAP) is calculated by plotting (entire) precision at various levels of (entire recall) and calculating the area under the curve.

The experiment was run using both decay functions: negative exponential and power. For the negative exponential model we set \(b = -0.001\) where distance is in meters; for the power model we set \(b = -0.5\). Various values were tried and these values gave good performance in terms of MAP over all the data sets. Each generative model may use either the uniform or Casablanca prior. This gives four variants. We include also the simplistic model described in Sec. 1 for comparison purposes. Note that for all the models except Burglary B the leave-one-out method was used – that for each crime being considered, all other crimes were used to estimate the home base of each offender. For Burglary B there was no need to estimate the home base since it was known.

5.3. Results for direction-free model

Table 2 shows the average precision measures for all the crime sets. Average entire precision is, roughly speaking, the area under the curve (in Fig. 4, for example). We can see that for all 6 datasets, the Generative Decay Model outperforms the simplistic model and that the differences between both forms of the decay function is small. The Casablanca prior outperforms the uniform prior in all but one case. Statistical significance testing, using the Friedman’s test, confirm our findings. Looking more specifically at two cases, Fig. 3 shows the precision and recall graph for Burglary A using the exponential function and Fig. 4 shows Criminal Damage using the power function. We can see that for lower levels of recall the precision is markedly higher. Therefore when the probability assigned to the prime suspect is high in absolute terms there will be much higher confidence in the accuracy of the prediction of who the culprit is.
Fig. 3. Entire precision and recall graph for negative exponential — Burglary A.

Fig. 4. Entire precision and recall graph for power model — Criminal Damage.

Table 2. Average entire precision for different crime sets.

<table>
<thead>
<tr>
<th>Crime Type</th>
<th>Neg. Exponential</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theft from Vehicles</td>
<td>0.286</td>
<td>0.200</td>
</tr>
<tr>
<td>Criminal Damage</td>
<td>0.472</td>
<td>0.399</td>
</tr>
<tr>
<td>Damage to Vehicles</td>
<td>0.863</td>
<td>0.881</td>
</tr>
<tr>
<td>Burglary A</td>
<td>0.114</td>
<td>0.075</td>
</tr>
<tr>
<td>Robbery</td>
<td>0.395</td>
<td>0.364</td>
</tr>
<tr>
<td>Burglary B</td>
<td>0.217</td>
<td>0.150</td>
</tr>
</tbody>
</table>
For the above results, the value of the $b$ parameter was set to a typical value for all data sets. We now attempt varying the parameter to determine how it affects the model. This has two uses: firstly, the value of the parameter can be selected to optimise the performance of the model. Secondly, we can attempt to estimate the value of $b$ to determine the shape of the decay function and in particular to establish that there is actual decay since a non-negative value of the parameter for each model would indicate no decay. Figures 5 and 6 show the varying of the parameter for burglary A and in both cases a clear maximum can be seen. It is worth noting that the accuracy of the model is not greatly influenced by the choice of parameter if we compare the range of average entire precision with the simplistic model. The graphs also provide evidence that there is actual decay since a value of $b = 0$ implies a flat decay function in both models i.e. there is no decay at all.
5.4. Results for directional model

By applying the test for directional Preference as defined in Subsec. 4.1 applied to the dataset Burglary B, the significance level is better than 0.001. We now compare the directional model with the unidirectional model. Note that when the parameter $w=0$ the Directional Model degenerates to the direction-free one. Thus we apply the model to Burglary Set B with $w=0, 0.5$ and $1$ using the Casablanca prior and the power function; results for exponential function are very similar. Figure 7 shows the entire precision and recall graph for all three values of $w$ for the power function. The average entire precision values are shown in Table 3 for both the power and exponential functions. This shows that adding the directional effect improves the model slightly.

6. Conclusions and Future Work

We have brought together both decay functions and a generative probabilistic model for suspect prioritisation to produce a predictive model which works more effectively than simply ranking distances from the offender’s base to the crime scene. To achieve this we require decay functions which are actual probability density functions and two functional forms fit the necessary criteria: negative exponential and power function. When applied to actual crime data there is no evidence that one model works consistently better than the other, although the power function does have the advantage that its parameter is dimensionless and thus does not depend...
on the units used to measure distance. The choice of parameter does affect performance but there appears to be a stable range which gives a good performance. We can use the crime data to estimate heuristically the single parameter of the model and show that the optimal parameter for each model is consistent with decay occurring as distance from the offender’s base increases. Finally, when we consider both the distance travelled and also preferred direction the model makes slightly more accurate predictions.

Finally, while in the work presented in this paper the models’ parameters were set manually, it is possible to estimate them from the data. For example, for the exponential model the parameters could be easily estimated by maximum likelihood from the data. In addition, the entire model selection could be carried out in a more principled way, by carrying out goodness of fit test in order to check whether or not the data come from the relative model, for instance the exponential model. We plan to implement these improvements in the next implementation of the models.

References