A Decision Model for Spatial Site Selection by Criminals: A Foundation for Law Enforcement Decision Support

Yifei Xue Donald E. Brown
yx8d@virginia.edu brown@virginia.edu
Department of Systems and Information Engineering
University of Virginia, Charlottesville, VA 22904, USA.

Abstract: Crime analysis uses past crime data to predict future crime locations and times. Typically this analysis relies on hot spots models that show clusters of criminal events based on past locations of these events. It does not consider the decision making processes of criminals as human initiated events susceptible to analysis using spatial choice models. This paper analyzes criminal incidents as spatial choice processes. Spatial choice analysis can be used to discover the distribution of people's behaviors in space and time. Two adjusted spatial choice models that include models of decision making processes are presented. The comparison results show that adjusted spatial choice models provide efficient and accurate predictions of future crime patterns and can be used as the basis for a law enforcement decision support system. This paper also extends spatial choice modeling to include the class of problems where the decision makers' preferences are derived indirectly through incident reports rather than directly through survey instruments.

Key words: Crime analysis, spatial choice, alternative sampling, feature selection, preference mining.

1. Introduction

Crime analysis involves exploiting data about crimes to enable law enforcement to better apprehend criminals and prevent crimes. Data used by crime analysts includes the time and locations of crimes and a variety of characteristics, such as methods of entry and items stolen, that vary with the type of crime. Crime analysts use these data with methodologies like aggregate crime rate analysis, hot spots, and space-time point process modeling to analyze and predict the spatial patterns of crimes.

Aggregate crime rate analysis uses sample units, such as neighborhoods, cities, and schools, to explain the variation in crime rates across those units. The statistical basis of this analytic approach is well established. Many regression methods, especially Poisson regressions are well studied and broadly used [6], [10], [16].

Hot spots models use past crime data to identify unusual clusters of criminal incidents within a well defined region. These clusters are commonly referred to as "hot spots" representing areas that contain unusual amounts of crimes. The term "hot spots" has become part of the crime analysis lexicon and has received a lot of attention [11]. Using the coordinates given by longitude and latitude, the locations of crime incidents are displayed and the prediction of possible future crimes is also indicated. One of the most promising methods by which to locate hot spots is with a density surface. Many results have been obtained for mapping and comparison of different crime density estimates [4], [7].

A new method based on space-time point process models is different from hot spots mapping. It predicts future crime locations based on the features of past crime data. The space-time prediction problem is formulated as an estimation of the transition density of a stochastic process. Feature space analysis enables the model to predict the potential locations of future events. Using a criminal incident database from the city of Richmond Virginia, Brown et al. demonstrate the performance improvement of adding dimensions of preferences discovered through feature space analysis [5].

However, none of these methods, including space-time point process models explicitly consider the decision making processes of the criminals. Criminal incidents, like many other human initiated events, can be described as a spatial choice procedure. They are linked to decision making processes indicating preferences that individuals have for specific sites in terms of certain spatial attributes. The analysis of spatial choice provides a methodology suited to problems related to site selection in geographical space and is concerned with human decision making processes.

Spatial choice studies come from McFadden's discrete choice theorem [14]. Discrete choice models are used for the analysis and prediction of people's choices when they face multiple alternatives. With the development of geographical information systems, spatial choice approaches have made great progress in characterizing human decision making in a spatial environment. All of this work builds on McFadden's initial results. Some research fields, like consumer destination selection [8], [17], travel mode analysis [15], [2] and recreational demand models [18] are characterized by an analysis of spatial patterns that indicate behavioral rules of a large number of individuals. These behaviors are all supported by informed decision making processes.

The spatial analysis of crimes differs from these other spatial choice models in important ways. The number of possible alternatives is very large and the subjects of our analysis (criminals) are not available for surveys as they are in consumer choice problems. We can only mine their preferences from past crime incidents and spatial features. Based on the spatial choice theories, we have developed a method to predict future crimes with the use of data from collected incident reports and analysis of offenders' decision making processes.

In the rest of this paper, we first define spatial choice process in Section 2 and describe their use in crime analysis in Section 3. Then two adjusted spatial choice models are provided in Section 4. In Section 5, the two models are applied to crime spatial analysis using real data for the prediction of future criminal incidents. Comparison results of the new models are also reported in Section 5 and Section 6 has results and conclusions.

2. Spatial choice theories

2.1 Problem statement

Data items for spatial or crime analysis have two components: a spatial component and an attribute component. Spatial data can be represented by a vector $\{Q, S, k\}$. Q is the spatial component, which is discrete and indexes all alternatives by an ordered pair of coordinates $\{x, y\}$. S is the attribute component associated with a given alternative, which indicates the S

different attributes of the spatial alternative, $S = \{s_1, s_2, ..., s_S\}$. $k : Q \to S$ is a mapping function specifying the link between observed alternatives and attribute values.

In spatial decision analysis, the decision process is represented by a vector $\{Q, S, k, A, D, u, P\}$. The set A is a subset of Q indicating the finite choices available to individuals. $A = \{ a_1, a_2, ..., a_N \}$ represents N available alternatives for decision makers to choose. For spatial analysis, N can be a very large number. D is the universe of individuals who make choices over the available alternative set A. Each individual makes choices based on decision processes. u is a utility function mapping the preferences from individuals d over the alternative set Q to a utility value U. For a certain individual d, if choice set $A_d = \{ a_1, a_2, ..., a_N \}$ and $A'_d = \{ a'_1, a'_2, ..., a'_N \}$ have same attribute values $S_A = k(A_d) = k(A'_d)$, then the choice sets will have same utility values $u(A_d) = u(A'_d)$. According to the assumptions of decision theory, individuals make choices that maximize their utility U. The choice probability is determined by utility maximization. The probability that an individual d from D will choose alternative a_i from available choice set A can be specified as $P\{ a_i \mid A_d, d \}$. It is derived from the choice process $\{Q, S, k, A, D, u, P\}$. The probability $P\{ a_i \mid A_d, d \}$ is a mapping based on the preferences of individual d and the attributes of alternatives in set A_d. The mapping can be stated as $P : A \times S \times D \to (0,1)$, or indicated by the utility based function $P\{ a_i \mid A_d, d \} = P\{ u(a_i) \geq u(a_j) | d \}$, for all $a_j \in A_d$.

2.2 Random utility maximization

Spatial choice is a decision process and can be described by a utility function. According to Keeney and Raiffa [12], an individual's utility is a function of the features or attribute values of alternatives. A basic approach to the mathematical theory of individual preferences derives from microeconomic consumer theory [1]. An individual consumer chooses a consumption bundle or alternative $Q= \{q_1, q_2,...,q_L\}$, where $q_1,q_2,...,q_L$ are quantities of commodities and services. $p_1, p_2,...,p_L$ are prices of commodities and services. For fixed income, I, the budget constraint of the individual is $\sum_{l=1}^{L} p_l q_l \leq I$. The consumer is assumed to have preferences over alternative consumption bundles. $Q^i \geq Q^j$ means consumption bundle Q^i is at least as good as consumption bundle Q^j. For any alternative bundles Q^i, Q^j and Q^k, if $Q^i \geq Q^j$ and $Q^j \geq Q^k$, then $Q^i \geq Q^k$. Based on these assumptions, the ordinal utility function is defined as $U(Q) = U(q_1, q_2,...,q_L) = \beta_0 + \sum_l \beta_l q_l$. If $Q^i \geq Q^j$, then $U(Q^i) \geq U(Q^j)$. The β 's represent the preferences of the consumer. The selection of the most preferred bundle can be stated as an optimization problem subject to the budget constraint. $max\, U(Q)\ \ s.t. \sum_{l=1}^{L} p_l q_l \leq I$.

Solving this optimization problem provides us with the consumer's preferences for his choice behaviors.

This economic consumer theory is developed without any assumptions about the nature of the alternatives. For discrete choice analysis, the overall choice set consists of discrete alternatives Q's. The values of $q_1, q_2,...,q_L$ are replaced with attribute values of alternatives and characteristics of decision makers, $X = (S, D)$. The choice set Q is finite and determined by all reality restrictions like budget and time. The consumer's choice is an alternative from the finite choice set that gives the highest utility. Solving the optimization problem, we can infer the information of consumer's preferences from their choice behaviors.

When a consumers' utility function is deterministic, his or her preferences can be easily estimated. However, most individuals' utility functions are not deterministic. Their choices and preferences are frequently inconsistent. To allow for this inconsistency, random utility theory was introduced into consumer theory by Manski and McFadden [13]. Random utility theory treats the unknown utility functions as random variables also because the analysts are unable to observe all attributes or consider variations of individuals. The utility of alternative a_i to individual d can be divided into two parts $U_{id} = V(d,s_i) + \varepsilon(d,s_i)$. $V(d,s_i) = \sum_l \beta_l^i x_l^i$ is the deterministic part of the utility function and expressed as a linear additive function of all attributes. $x_l^i \in X = (S,D)$ represents the lth value of the combination of attribute values s_i and characteristics of individual d. $\varepsilon(d,s_i)$ is the error term of utility function indicating unobservable components of the utility function. With the random utility function, we are unable to identify the alternative that has the highest utility for consumers. We only know that the probability of a chosen alternative with highest utility $P(a_i|A_d) = P(U_{id} > U_{jd}, all\, a_j \in A_d)$ will be the biggest among all alternatives in choice set. Based on certain assumptions of the error term, the probability that an alternative has the highest utility can be estimated as

$$
\begin{aligned}
P\{a_i \mid A_d, d\} &= P\{V(d,s_i) + \varepsilon(d,s_i) \ge V(d,s_j) + \varepsilon(d,s_j), \textit{for all } j \ne i, a_j \in A_d\} \\
&= P\{\varepsilon(d,s_j) - \varepsilon(d,s_i) \le V(d,s_i) - V(d,s_j), \textit{for all } j \ne i, a_j \in A_d\}
\end{aligned} \tag{1}
$$

The stochastic part $\varepsilon(d,s_i)$ is usually assumed to be independently and identically distributed with Weibull distribution [14]. The specific structure of the discrete choice model can be stated as

$$
P(a_i \mid A_d, d) = \exp V(d,s_i) / \sum_{a_j \in A_d} \exp V(d,s_j) \tag{2}
$$

The probability of individual *d*'s choice for a_i is predicted by equation (2). The preferences or parameters of the utility function β's are estimated by methods of maximum likelihood estimation [1].

3. Crime analysis

Crime analysis uses past crime data to analyze and predict the behavior rules of criminals in space. Spatial criminal incident prediction is usually carried out within a specified geographic region. The possible targets within the region are spatial alternatives of criminals' choices. With the following analysis of crimes, we will incorporate the spatial choice model into a methodology for crime analysis and use the result as the basis for new models of spatial choice presented in Section 4.

3.1 Background of the crime analysis project

The data for model estimation comes from ReCap (the Regional Crime Analysis Program). ReCap is an interactive shared information and decision support system that uses databases, geographic information system (GIS), and statistical tools to analyze, predict and display future crime patterns.

Our crime analysis is based on criminal incidents between July 14 1997 and August 10 1997 in the city of Richmond Virginia. We use residential "Breaking and Entering" (B & E) crime incidents for model estimation and validation. Using the first four weeks of crime incidents as the training dataset, we get locations of criminal incidents on a geographic map. The sub regions shown in figure 1 are census block groups, which are the smallest areas for that census counts are recorded.

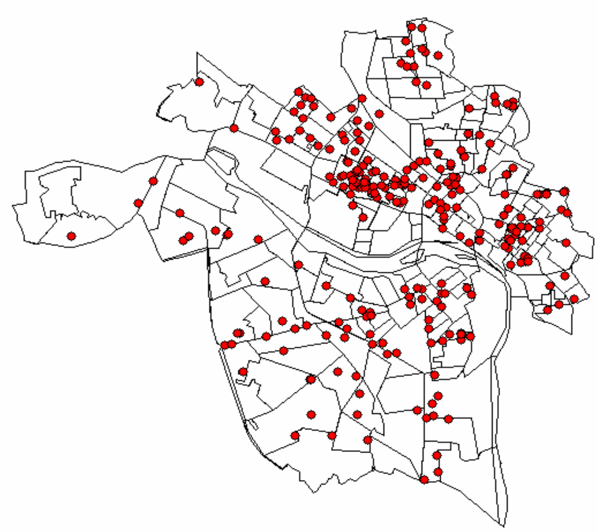

Figure 1. Breaking and Entering criminal incidents

between July 14, 1997 and August 10, 1997 in Richmond, Virginia.

The analysis of B & E is related to locations of households in a city. However, it is difficult to represent all locations of individual houses in even a modest sized city, such as Richmond. Therefore, we aggregate alternatives using a regular grid of 2517 points, which is assumed to be fine enough to represent all spatial alternatives within this area. The features of each spatial alternative come from the combination of census data and calculated distance values. All features have been shown to be related to the decision process of criminals in this area [5].

3.2 Feature selection by similarities

Since the attributes of spatial alternatives come from census data and calculated distance values, it is possible that some values of these attributes are correlated. Using the calculated correlation value as similarity, we perform a hierarchical clustering algorithm on all features used for spatial choice. The clustering of features for observed spatial alternatives is shown in Figure 2.

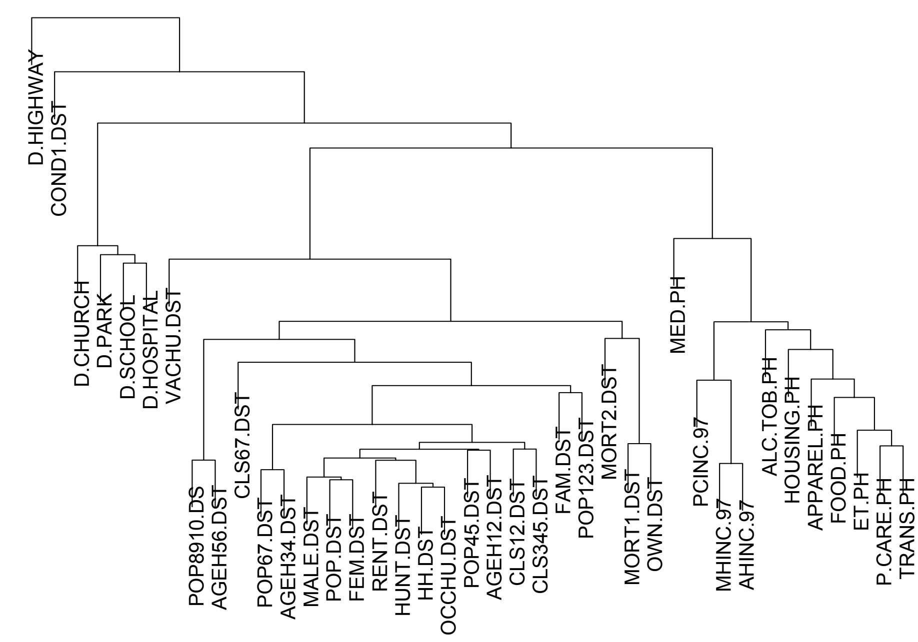

Figure 2. Clusters of features for observed spatial alternatives

From this result, we divide the features into five clusters. Each cluster includes features with correlated values. After checking the distribution of feature values, we find that some features are almost uniform and therefore do not represent good features for our analysis. In order to be consistent with Brown et al. [5], we pick same features that they used: distance to highway (D.HIGHWAY), family density per unit area (FAM.DENSITY), personal care expenditure per household (P.CARE.PH) and an extra feature distance to hospital (D.HOSPITAL). These are key features used in our models.

4. Model development

4.1 Crime processes as spatial choices

Spatial choice is a process that individuals use to choose a specific site in space to meet their objectives. Criminals choose possible sites in space to commit crimes. Their choices show certain patterns over a geographic region. The geographical sites form a universal spatial alternative set A. Individuals will make a selection from the choice set A. The spatial alternatives in choice set A have some specific characteristics, which make the spatial choice process different from other choice processes. First, the spatial choice process is a discrete process. The number of alternatives, N will be finite and large. For some cases, N could be very large. Second, the spatial alternatives have relatively stable positions during the choice process.

For spatial choice problems, the number of alternatives is very large. Individuals are unable to evaluate all spatial alternatives before they make their selections that have the highest personal utility. They can only compare part of the universal choice set. This can be stated as a sub-optimal or locally optimal problem. According to the framework of Fotheringham et al. [9] concerning individuals' hierarchical information processing, individuals make spatial choices from alternatives with features that they have evaluated. For individual *d*, the choice set will be $A_d \subseteq A$, is the choice set that individual *d* really considered. The choices that individual *d* makes will most probably have the highest utility among all alternatives in choice set A_d. Unlike traditional discrete choice theory, the real choice set A_d that is first evaluated by the individual is not known to analysts. Some methods are proposed to identify or estimate the probability that an alternative a_i is considered by individual *d*, $P(a_i \in A_d)$.

During the identification of an individual's choice set, two factors are considered in his spatial choice procedure: *i*) the utility of alternative a_i to individual *d* and *ii)* the possibility that an alternative is available or considered by individual *d*. Since the number of alternatives in a spatial choice problem is very large, it is possible that some alternatives can have higher utility values but are never considered. If we assume that the two factors are equally important for the individuals' choice, then the product of $P(a_i \in A_d)$ and the utility of alternative a_i to individual *d*, U_{id}, can give a better estimate of the likelihood of choice. Then the probability that individual *d* chooses alternative a_i from choice set A_d can be stated as $P(P(a_i \in A_d) \cdot U_{id} > P(a_j \in A_d) \cdot U_{jd}, all\, a_j \in A_d)$. Under the extreme distribution assumption, we get the spatial site selection model using same derivation as McFadden [14] and Fortheringham [8].

$$P(a_i \mid A_d, d) = exp(V(d, s_i)) \cdot P(a_i \in A_d) / \left(\sum_{a_j \in A} exp(V(d, s_j)) \cdot P(a_j \in A_d) \right). \tag{3}$$

This model is a multinomial logit model where each alternative's observable utility is weighted by the probability that alternatives are being evaluated.

4.2 Specification of priori

We assume that hierarchical information processing takes place before an individual's spatial choice. An individual will first evaluate a set of alternatives and only alternatives within the evaluated set will be selected. We need to define a probability that an individual d will evaluate alternatives a_i: $P(a_i \in A_d)$.

For spatial analysis, such as crime analysis, it is not easy to know an individuals' preferences and characteristics. Here we assume that the characteristics of all individuals $d \in D$ are same. The pre-evaluated choice set A_d is also the same to all individuals. We use M to represent the same choice set for all individuals. Then the spatial site selection model changes to

$$P(a_i \mid M, d) = exp(V(d, s_i)) \cdot P(a_i \in M) / \left(\sum_{a_j \in A} exp(V(d, s_j)) \cdot P(a_j \in M) \right) \tag{4}$$

Hot spots mapping models and space-time prediction models are used to indicate the possibility of alternative a_i being evaluated by criminals $P(a_i \in M)$.

Hot spots are typically located by density estimation, such as kernel density estimation or mixture model. The estimated density surface maps the past criminal incidents in space. The estimated density surface gives the probability that an alternative will be considered by criminals in the future.

In order to reveal the true preferences of criminals, Brown et al. provide a new point process prediction model. In this model, they identify individuals' choices by analyzing their preferences for all alternatives. Suppose we have l measurable features $f_1, f_2, \ldots, f_l$ that are known or believed to be relevant to the occurrence of criminal incidents. Then the hyperspace formed by these l features is called a feature space. However, it is difficult to know exactly which features are actually considered by criminals. Brown et al. find the smallest feature subset, called key feature set or key feature space, with a feature selection process. The past criminal incidents indicate clear patterns in the key feature space. Then using density transition estimation in key feature space, the probabilities of alternatives that could be considered by criminals are estimated.

4.3 Sampling of alternatives

The number of spatial alternatives for problems such as crime analysis is very large. This makes the data preparation and computation time prohibitively expensive. To handle this problem, we use a subset of the alternatives, including all chosen alternatives and a sample of non-chosen alternatives to get consistent estimation of the choice model, which is called importance sampling by Ben-Akiva and Lerman [1]. Sampling alternatives is an easily applied technique to reduce the computational burden involved in the model estimation process.

Suppose there are K observations in the estimation procedure. For each observation a_k, we randomly assigned j_k spatial alternatives from choice set M to form a choice set C_k, which is far less than the total number of spatial alternatives N. All choice sets C_k, $k=1 \ldots K$ form the estimation choice set C. $\Pi(C|a_i)$ indicates the conditional probability of constructing sampled choice set C for observation a_i. Using Bayes' theorem, the conditional probability of alternative a_i being chosen given a sample of alternatives C is

$$\Pi(a_i|C) = \frac{\Pi(C|a_i)P(a_i|M,d)}{\sum_{a_j \in C} \Pi(C|a_j)P(a_j|M,d)} \tag{5}$$

$P(a_i|M,d)$ is the probability that individual d will choose alternative a_i from the choice set M. Replacing equation (4) into equation (5), we obtain the conditional probability of alternative a_i being chosen through the sampled choice set C as

$$\Pi(a_i|C) = \frac{exp[V(d,s_i) + ln\,P(a_i \in M) + ln\,\Pi(C|a_i)]}{\sum_{a_j \in C} exp[V(d,s_j) + ln\,P(a_j \in M) + ln\,\Pi(C|a_j)]} \tag{6}$$

Using Ben-Akiva and Lerman's importance sampling methodology, we perform K independent draws from all alternatives of A and then add the observed choice a_k to construct the sampled choice set C. The sampling prior that the jth alternative will be chosen is q_j. The resulting sample of alternatives is characterized by the probability distribution

$$\Pi(C|a_i) = \prod_{\substack{a_j \in C \\ a_j \neq a_i}} q_i \prod_{a_j \notin C} (1 - q_j) = \frac{1}{q_i} Q(C) \tag{7}$$

where $Q(C) = \prod_{a_j \in C} q_j \prod_{a_j \notin C} (1 - q_j)$ is the unconditional selection probability and is independent of the chosen alternative. Substituting equation (7) into (6), we derive

$$\Pi(a_i|C) = \frac{exp[V(d,s_i) + ln\,P(a_i \in M) + ln\frac{1}{q_i}Q(C)]}{\sum_{a_j \in C} exp[V(d,s_j) + ln\,P(a_j \in M) + ln\frac{1}{q_j}Q(C)]}$$

$$= \frac{exp[V(d,s_i) + ln\,P(a_i \in M) + ln\frac{1}{q_i}]}{\sum_{a_j \in C} exp[V(d,s_j) + ln\,P(a_j \in M) + ln\frac{1}{q_j}]} \tag{8}$$

Before the sampling process, we specify the prior of being chosen for the *jth* alternative. Since the importance sampling strategy assigns a higher prior to an alternative that is most likely to be chosen, we take the probability that an individual will evaluate the *jth* alternative as $p(a_j \in M)$ and refer to it as the sampling prior q_j. Our spatial selection model becomes

$$\begin{aligned}\Pi(a_i|C) &= \frac{exp[V(d,s_i) + ln\,P(a_i \in M) - ln\,P(a_i \in M)]}{\sum_{a_j \in C} exp[V(d,s_j) + ln\,P(a_j \in M) - ln\,P(a_j \in M)]} \\ &= \frac{exp\,V(d,s_i)}{\sum_{a_j \in C} exp\,V(d,s_j)}\end{aligned} \quad . \tag{9}$$

This is the simple multinomial logit model. The estimation of this model gives us a consistent estimation to the spatial choice $P(a_i|M,d)$ in equation (4).

5. Crime analysis results

5.1 Model estimation and prediction

Based on above analysis, we next consider the model estimation and prediction step. Two methods of specifying priors are proposed as following.

Using a hot spots model, the prior $P(a_i \in M)$ is estimated with kernel density estimation.

$$P(a_i \in M) = \frac{1}{K}\sum_{k=1}^{K} T(\frac{x_i - x_k}{h_1}, \frac{y_i - y_k}{h_2}) \tag{10}$$

where x_i, y_i are coordinates of spatial alternative a_i. K is the total number of observations. T is a function to specify the kernel estimator. In this work we use Gaussian function: $T(\frac{x_i - x_k}{h_1}, \frac{y_i - y_k}{h_2}) = \frac{1}{2\pi h_1 h_2} exp[-\frac{1}{2}((\frac{x_i - x_k}{h_1})^2 + (\frac{y_i - y_k}{h_2})^2)]$. h's are bandwidths used in the kernel estimation. The change of bandwidth will influence the effect of density estimation. The choice of bandwidth is important and literature in this area offers a great deal of discussion on the choice of bandwidth. We use a recommended bandwidth selection from Bowman and Azzalini [3], $h_l = \left(\frac{4}{(l+2)\cdot K}\right)^{1/(L+2)} \times \sigma_l$ for lth dimension, $l=1, 2$. L is the total number of dimensions for density estimation. σ_l is the standard deviation of observed data on dimension l. The estimated model is called space adjusted spatial choice model.

In accordance with our earlier discussion of spatial feature selection, we use families per unit area, personal care expenditures per household, distance to highway and distance to hospital as the key features. We next apply density estimation to this key feature space and obtain the prior evaluation probability

$$P(a_i \in M) = \frac{1}{K}\sum_{k=1}^{K} T\left(\frac{x_i^1 - x_k^1}{h_1}, \frac{x_i^2 - x_k^2}{h_2}, \frac{x_i^3 - x_k^3}{h_3}, \frac{x_i^4 - x_k^4}{h_4}\right) \tag{11}$$

where x_i^1, x_i^2, x_i^3 and x_i^4 are key features of spatial alternative a_i. K is the total number of observations, and T is a Gaussian function to specify the kernel estimator. h's are bandwidths. The estimated model here is called key feature adjusted spatial choice model.

Using the training data set of B & E incidents between July 14, 1997 and August 10, 1997, we obtain the estimation of the two adjusted spatial choice models. With the estimation results, we make predictions of future crime locations. Two testing data sets are specified for model evaluations and comparisons. One testing data set contains incidents between August 11, 1997 and August 17, 1997. The other testing data set contains incidents between August 11, 1997 and August 24, 1997. The prediction and testing incidents for hot spots model and our two adjusted spatial choice models are displayed in Figure 3-5.

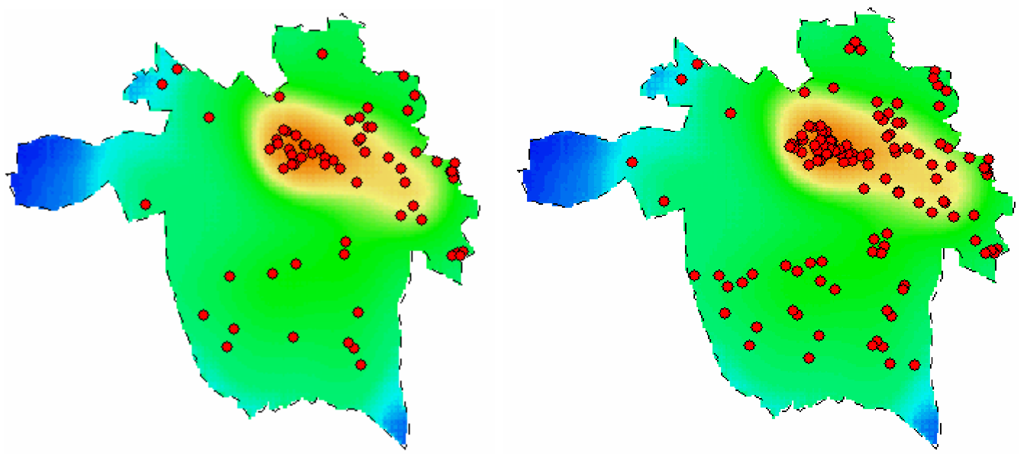

Figure 3. Prediction of hot spots model with crime incidents

from 08/11/97 to 08/17/97 and 08/11/97 to 08/24/97

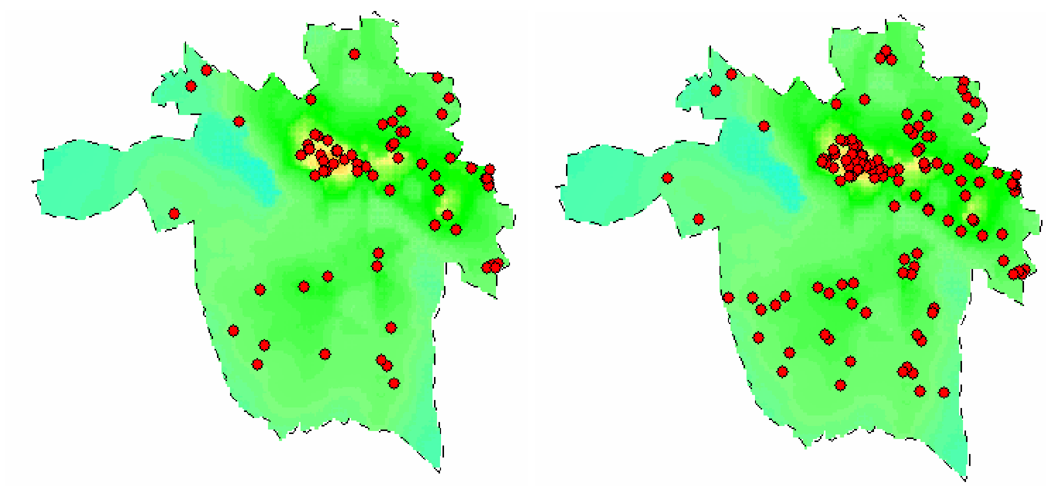

Figure 4. Prediction of space adjusted spatial choice model with crime incidents

from 08/11/97 to 08/17/97 and 08/11/97 to 08/24/97

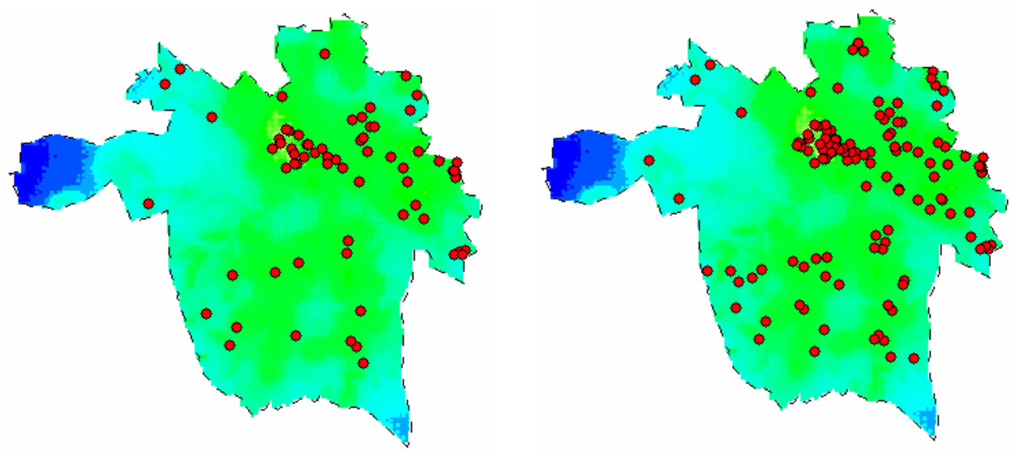

Figure 5. Prediction of key feature adjusted spatial choice model and crime incidents from 08/11/97 to 08/17/97 and 08/11/97 to 08/24/97

5.2 Model comparisons

To compare different models, we standardize the predictions of the three models. The hypothesis is that for the population of all future crime incidents, the proposed model will statistically outperform the comparison model.

Assume that the testing data set contains m incidents that occurred at locations $a_1, a_2, \ldots, a_m$, respectively. For incident a_i, let the predicted probability given by the proposed model be $p^p_{a_i}$ and that given by the comparison model be $p^c_{a_i}$. The hypothesis test is built around μ which denotes the mean of the difference between predicted probabilities given by the proposed model and that given by the comparison model.

Assume that the proposed model have better prediction than the comparison model for all future crimes. Then the null hypothesis is that the predicted probability difference μ between the two models for all future crime incident locations is less than or equal to 0. The alternative hypothesis is that the predicted probability for proposed model is significantly better than the comparison model. We perform the hypothesis test as follows:

$$\begin{aligned} &\mathrm{H}_0\colon \mu \le 0\,, \\ &\mathrm{H}_a\colon \mu > 0\,. \end{aligned} \tag{12}$$

Using testing data set with *m* crime incidents, we obtain the estimated probability difference $\hat{\mu}$.

$$\hat{\mu} = (1/m)\sum_{i=1}^{m}\left(p_{a_i}^{p} - p_{a_i}^{c}\right) \tag{13}$$

The standard deviation of the difference, $q_{a_i} = p_{a_i}^{p} - p_{a_i}^{c}$ is estimated by

$$\hat{\sigma} = (1/(m-1))\sum_{i=1}^{m}\left(q_{a_i} - \hat{\mu}\right)^2 . \tag{14}$$

The results of these tests are shown in table 1. In the testing results, “Mean” and “Std. Dev.” stand for $\hat{\mu}$ and $\hat{\sigma}$, respectively. *p*-value indicates the type I error for rejecting the null hypothesis.

Table 1. The model comparison results

Testing data set 1 (08/11/97 - 08/17/97)

	Mean	Std. Dev	z-Statistic	p-Value
Space Adjusted vs. Hot Spot	4.890×10^{-4}	7.902×10^{-4}	5.251	<0.0001
Key Feature Adjusted vs. Hot Spot	8.963×10^{-4}	1.286×10^{-4}	5.913	<0.0001
Key Feature Adjusted vs. Space Adjusted	4.072×10^{-4}	1.020×10^{-4}	3.387	<0.0001

Testing data set 2 (08/11/97 - 08/24/97)

	Mean	Std. Dev	z-Statistic	p-Value
Space Adjusted vs. Hot Spot	5.308×10^{-4}	7.972×10^{-4}	7.5336	<0.0001
Key Feature Adjusted vs. Hot Spot	9.285×10^{-4}	1.231×10^{-3}	8.5328	<0.0001
Key Feature Adjusted vs. Space Adjusted	3.977×10^{-4}	9.985×10^{-4}	4.5056	<0.0001

The comparison results above indicate that the two adjusted spatial choice models dramatically outperform the hot spots model. Furthermore, the key feature space adjusted spatial choice model improves the prediction results over space adjusted spatial choice model significantly. According to previous studies [5], the key feature space density estimation gives a better prediction of the prior for each alternative to be evaluated than space density estimation does. Based on the specification of priors in feature space over all alternatives in choice set, we provide a more efficient and accurate method for predictions of criminal's future spatial choices. This result has clear implications for other applications.

6. Conclusion

In this paper, two models for criminal site selection, the space adjusted spatial choice model and the key feature adjusted spatial choice model are presented by modifying a traditional discrete choice model. Using real crime incidents, we show that the two adjusted spatial choice models can reveal the preferences of unknown criminals in space and give more accurate predictions of spatial patterns of future crimes. These methods perform significantly better than the hot spots method currently used by most police agencies. The presented models also provide a new approach to crime analysis and can serve as the basis for decision support. We plan to incorporate these models into the ReCAP decision support system, which is currently in use in several law enforcement agencies.

References

[1] M. Ben-Akiva and S. Lerman, Discrete Choice Analysis, Theory and Application to Travel Demand. Cambridge, MA: MIT Press, 1985.

[2] C. Bhat, "Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling," Transportation Science, Vol. 34, No. 2, pp. 228-238, May 2000.

[3] A. Bowman and A. Azzalini, Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations. Oxford NY: Oxford University Press, 1997.

[4] S. Bremer, “An exploration of the methods for detecting hot spots and changes in hot spot locations,” M. S. Thesis, University of Virginia, Charlottesville, VA, 2000.

[5] D. Brown, H. Liu and Y. Xue, “Mining preferences from spatial-temporal data,” Proceedings of SIAM Conference, 2001.

[6] A. Cameron and P. Trivedi, Regression Analysis of Count Data. Cambridge: Cambridge University Press, 1998.

[7] J. Dalton, “Bandwidth selection for Kernel Density Estimation of geographic point processes,” M. S. Thesis, University of Virginia, Charlottesville, VA, 1999.

[8] S. Fotheringham “Consumer store choice and choice set definition,” Marketing Science, pp. 299-310, summer, 1988.

[9] S. Fotheringham, C. Brunsdon and M. Charlton, Quantitative Geography. Thousand Oaks, CA: SAGE, 2000.

[10] W. Gardner, E. Mulvey, and E. Shaw, “Regression analyses of counts and rates: Poisson, over dispersed Poisson, and negative binomial,” Psychol. Bull 118:392-405, 1995.

[11] K. Harries, Mapping Crime: Principle and Practice. DC: NIJ, CMRC, 1999.

[12] R. Keeney and H. Raiffa, Decisions with Multiple Objectives, Preferences and Value Tradeoffs. Cambridge, UK: Cambridge University Press, 1993.

[13] C. Manski and D. McFadden, Structural Models of Discrete Data. Sociological Methodology. Cambridge, MA: MIT Press, 1981.

[14] D. McFadden, Conditional Logit Analysis of Qualitative Choice Behavior. In Frontiers in Econometrics. P. Zarembka, ed. pp. 105-142, New York: Academic Press, 1973.

[15] D. McFadden and K. Train, "The goods/leisure tradeoff and disaggregate work trip mode choice models," Transportation Research, 12, pp. 349-353, 1978.

[16] D. Osgood, "Poisson-based regression analysis of aggregate crime rates," Journal of Quantitative Criminology, Vol. 16. No. 1, pp. 21-43, 2000.

[17] R. Rust and N. Donthu, "Capturing geographically localized misspecification error in retail store choice models," Journal of Marketing Research, Vol. XXXII, pp. 103-110, 1995.

[18] K, Train, "Recreational demand models with taste differences over people," Land Economics, 74, pp. 230-239, 1998.