Decision based spatial pattern analysis

Yifei Xue Donald E. Brown
yx8d@virginia.edu brown@virginia.edu
Department of Systems and Information Engineering
University of Virginia, Charlottesville, VA 22903, USA.

Abstract: The spatial pattern analysis can be used to discover the distribution of people's behaviors in space. Traditional spatial pattern analysis does not consider the decision making process of human activities. This paper presents two adjusted spatial choice models that include models of decision-making process. These new adjusted spatial choice models are applied to detect the spatial patterns of criminal incidents. The comparison results show that the adjusted spatial choice models provide efficient and accurate predictions of future crime patterns.

Key words: Spatial choice, alternative sampling, feature selection, preference mining.

1. Introduction

The rapid development in database technologies and data collection techniques has collected huge amounts of data in large databases. The growing data creates the necessity of knowledge and information discovery from data, which leads to an emerging field, data mining or knowledge discovery in databases (KDD). With the rapid increase of amount of spatial data collected from various applications, ranging from remote sensing, to geographical information systems (GIS), computer cartography, environmental assessment and planning, etc, it is also highly demanded for the applications of data mining or knowledge discovery to spatial data bases.

Spatial data are special comparing to the data in relational database. They carry topological and distance information and are often organized by spatial indexing structures. The spatial data for analysis could be classified as location data and attribute data. Location data consists purely of the locations at which a set of events occurred. It is also called event data, object data or a point process. Location data may be specific points, cell of a regular grid or irregular polygons. Attribute data consists of values, or attributes, associated with a set of locations.

Statistical spatial analyses have been the most common approaches for analyzing spatial data. Statistical spatial analyses include the analysis of point patterns, the analysis of spatially continuous data, the analysis of area data, and the analysis of spatial interaction data (Bailey, 1994). These analyses address the inherent stochastic nature of patterns and relationships. They are not deterministic and also not just simple descriptive analysis like data transformation and summarization as most common GIS functions. They are developed from different disciplines and have some general characteristics. These methods cover the scope of analytical problems, which commonly arise and benefit from realistic representation of space that GIS can provide. The most common ones are nearest neighbor methods and K-functions (Upton 1985, Ripley 1981), kernel smoothing methods (Silverman 1986), spatial autocorrelation and covariance modeling (Getis 1990), Geostatistical and spatial econometric modeling (Haining 1990, Cressie 1991), multivariate techniques (Oliver and Webster 1989), and spatial interaction models (Fortheringham 2000).

Crime analysis is typically an application of statistical spatial analysis. The geographic locations of past crimes can provide some information of the possible distributions of criminal behaviors. Based on past crime data, crime analysis uses methodologies like aggregate crime rate analysis, hot spots or space-time point process modeling to analyze and predict the spatial patterns crimes.

The aggregate crime rate analysis uses sample units, such as neighborhoods, cities, and schools, to explain the variation in crime rates across those units. This method is based on the spatial analysis of area data. Many regression methods, especially Poisson regressions are well studied and broadly used (Cameron and Trevedi, 1998; Gardner et al., 1995; Osgood, 2000). The aggregate crime rate analysis has strong assumptions about the coming out of crime incidents, which limit its prediction capabilities. The sample unit is still a big area and the predicted crime rates are not so helpful for law enforcement.

Hot spots models are applications of analysis of point patterns. It uses 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 amount of crimes. The term "hot spots" has become part of crime analysis lexicon and has received a lot of attention (Harris, 1999). Using the coordinate in longitude/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 comparisons of different crime density estimates (Bremer, 2000; Dalton 1999). The problem to hot spots model is that it only considers the spatial closeness of past crime incidents. All other spatial features are not considered in the model. The possible relationships between the crime incidents on certain spatial features are not detected.

The space-time point process model is an extension of point patterns analysis. It predicts the future crime locations based on the features or attributes attached with the location of past crimes. The space-time prediction problem is formulated as an estimation of the transition density of a stochastic process. The 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 and Liu demonstrates the performance improvement of adding the dimension of preference discovery through feature space analysis (Brown and Liu, 2001).

The crime analyses, including Brown and Liu's space-time point process model do not consider the decision making process. Criminal incidents, like many other human initiated events, can be described as a spatial choice procedure. It is linked with a decision-making process indicating preferences that individuals have for specific sites in terms of certain spatial attributes. The analysis of spatial choice introduced a methodology developed for 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 (McFadden, 1973; Ben-Akiva, 1985). Discrete choice model is used for analysis and prediction of people's choices when they face multiple alternatives. With the development of geographical information systems, spatial choice made great progress. Some research fields, like consumer destination selection (Fotheringham, 1988; Rust and Donthu, 1995), travel mode analysis (McFadden and Reid, 1975, Bhat, 1998); recreational demand models (Train, 1998) and crime pattern analysis are characterized by an analysis of spatial patterns that indicate certain behavior rules of a large number of individuals. Their behaviors are all supported by an informed decision making process.

The spatial analysis of crimes is different from other spatial choice analysis. The number of possible alternatives is very large and the preferences of decision makers or criminals are impossible to be revealed from survey or interview. We can only mine their preferences from past crime incidents and spatial features. Based on spatial choice theories, we will develop 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 the spatial pattern analysis problem as a spatial choice process. Then two adjusted spatial choice models are provided in section 3. In section 4, the two models are applied to criminal spatial pattern analysis for prediction of locations of future criminal incidents. Comparison results of the new models are also reported.

2. Problem statement

Data items for spatial or crime analyses 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 $S = \{s_1, s_2, ..., s_S\}$. $k : Q \rightarrow S$ is a mapping function specifying the observed attributes of alternatives.

In spatial decision analysis, the decision process is stated 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 the 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 $U = u(A_d) = u(A'_d)$. According to decision analysis assumptions, 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 $P\{a_i \mid A_d, d\}$. This probability is produced 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 $P : A \times S \times D \rightarrow [0,1]$, or indicated by the utility-based function $P\{a_i \mid A_d, d\} = P\{u(a_i) \geq u(a_j) \mid d\}$, for all $a_j \in A_d$.

3. Model development

3.1 Spatial choice patterns

Spatial choices describe the processes that individuals choose a specific site in space as their target. Their choices can show certain patterns in space. The geographical sites form a universal spatial alternative set *A*. Individuals will make a selection from the choice set *A*. The spatial alternatives in the choice set *A* have some specific characteristics, which make the spatial choice process different from the 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 with the highest 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 Fotheringham's framework of individuals' hierarchical information processing (Fotheringham, 2000), individuals make spatial choices from the alternatives they have evaluated. For individual *d*, the choice set will be $A_d \subseteq A$, which 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. Different from traditional discrete choice theory, the real choice set A_d is not clear to analysts. Some methods are proposed to identify or estimate the probability that an alternative a_i is considered by individual *d*, $P(a_i \in A_d)$.

After the identification of individuals' choice set, two factors are considered in people's spatial choice procedure: *i)* the utility of alternative *i* to individual *d* and *ii)* the possibility that alternative is available or considered by individual *d*. Since the number of alternatives in spatial choice problem is very large, it is possible that some alternatives can give higher utility value but they are never considered. If we assume the two factors are equally important to the individuals' choice decisions, 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 estimation of the possibility of choice. Then the probability that individual *d* chooses alternative a_i 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 assumptions, we get the spatial site selection model with same derivation as McFadden (McFadden, 1973; Fotheringham, 2000).

$$P(a_i \mid A_d, d) = exp(V(d, s_i)) \cdot p(a_i \in A_d) / \left(\sum_{j \in A} exp(V(d, s_j)) \cdot p(a_j \in A_d) \right). \quad (1)$$

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

3.2 Specification of priori

We assume that hierarchical information processing takes place before the individuals' spatial choices. Individuals will first evaluate clusters of alternatives and only alternatives within the selected clusters can be selected. We can either define the choice set A_d or give the probability that an individual will evaluate certain alternatives $P(a_i \in A_d)$.

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

$$P(a_i \mid A_d, d) = exp(V(d, s_i)) \cdot p(a_i \in M) / \left(\sum_{j \in A} exp(V(d, s_j)) \cdot p(a_j \in M) \right) \quad (2)$$

Hot spot mapping model and space-time prediction model are used to indicate the possibility of being evaluated by the criminals $P(a_i \in A_d)$. Hot spot mapping uses past crime data to identify unusual clusters of criminal activity within a region. 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 and Liu (2001) provide a new point process prediction model. In this model, they identify individuals' choices by analyzing preferences of individuals for all alternatives. Suppose we have p measurable features $f_1, f_2, ..., f_p$ that are known or believed to be relevant to the occurrence of the criminal incidents. Then the hyperspace formed by these p features is called a feature space. However, it is difficult to know exactly which features are actually considered by criminals. Brown and Liu (2001) found the smallest feature subset, called the 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, Brown and Liu (2001) estimate the probability of alternatives that could be considered by criminals.

3.3 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 the 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. Sampling alternatives is an easily applied technique for reducing the computational burden involved in the model estimation process (Ben-Akiva, 1985).

Suppose there are K observations in the estimation procedure. For each observation a_k, we randomly assigned j_k spatial alternatives from choice set A_d 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

$$\Pi(a_i|C) = \frac{\Pi(C|a_i)P(a_i|A_d,d)}{\sum_{j \in C} \Pi(C|a_j)P(a_j|A_d,d)} \tag{3}$$

$P(a_i|A_d,d)$ is the probability that individual d will choose alternative a_i from the choice set A_d. Replacing equation (2) into equation (3), we obtain the conditional probability of alternative a_i being chosen through the sampled choice set C as

$$\Pi(a_i|C) = \frac{exp[V(d,s_i) + ln\, p(a_i \in M) + ln\, \Pi(C|a_i)]}{\sum_{j \in C} exp[V(d,s_j) + ln\, p(a_j \in M) + ln\, \Pi(C|a_j)]} \tag{4}$$

Using Ben-Akiva's importance sampling methodology (1985), 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 priority that the *jth* alternative will be chosen is q_j. The resulting sample of alternatives is characterized by the probability distribution

$$\Pi(C|a_i) = \prod_{\substack{a_j \in C \\ a_j \neq a_i}} q_j \prod_{a_j \notin C} (1 - q_j) = \frac{1}{q_j} Q(C) \tag{5}$$

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. Placing equation (5) into (4), we derive a new equation

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

$$= \frac{exp[V(d,s_i) + ln\, p(a_i \in M) + ln \frac{1}{q_i}]}{\sum_{a_j \in C} exp[V(d,s_j) + ln\, p(a_j \in M) + ln \frac{1}{q_j}]} \tag{6}$$

Before the sampling process, we specify a priority of being chosen as the *jth* alternative. Since the importance sampling strategy assigns a higher priority to alternatives that are most likely to be chosen, we take the probability that an individual will evaluate certain alternative $p(a_i \in M)$ as the sampling priority q_j. Our spatial selection model becomes

$$
\begin{aligned}
\Pi(a_i|C) &= \frac{exp[V(d,s_i)+ln\,p(a_i \in M)-ln\,p(a_i \in M)]}{\sum_{a_j \in C} exp[V(d,s_j)+ln\,p(a_j \in M)-ln\,p(a_j \in M)]} \\
&= \frac{exp\,V(d,s_i)}{\sum_{a_j \in C} exp\,V(d,s_j)}
\end{aligned}
\quad . \tag{7}
$$

This is the simple multinomial logit model. The estimation of this model gives us a consistent estimation to the spatial choice $P(a_i|A_d,d)$ in equation (2).

4. Application to crime analysis

4.1 Background of the crime analysis project

The data for model estimation comes from ReCap (Regional Crime Analysis Program) system. The ReCap system 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 the crime incidents between July 14 1997 and August 17 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 the locations of the criminal incidents on the geographic map. The sub regions shown in figure 1 are block groups, which are the smallest areas for which census counts are recorded.

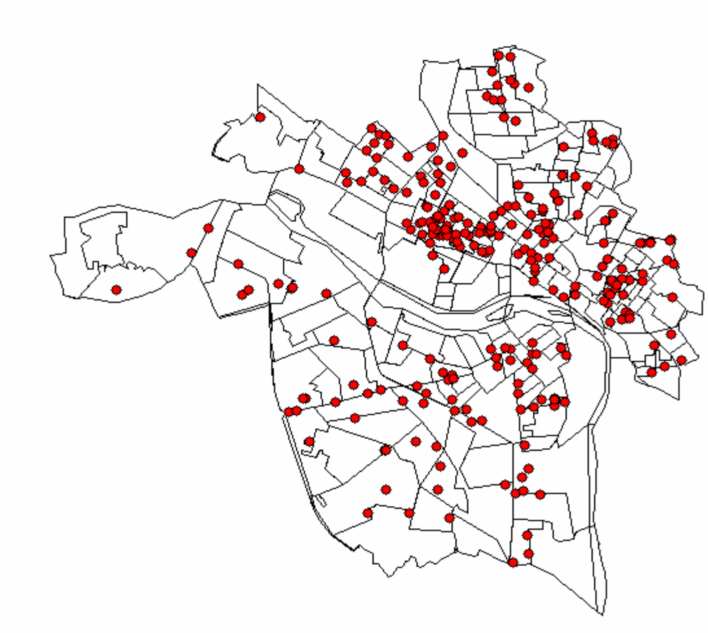


Figure 1. Breaking and Entering criminal incidents
between July 14, 1997 and August 10, 1997 in Richmond, Virginia.

The analysis of B & E is related with locations of households in a city. However, it is difficult to represent all locations of houses in a big city. We place over the Richmond area a regular grid of 2517 points, which is hopefully 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 are supposed to be related with the decision process of criminals (Brown and Liu, 2001).

4.2 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 make hierarchical clusters on all features of observed spatial choices. The clustering of features of observed spatial alternatives can be shown as below.

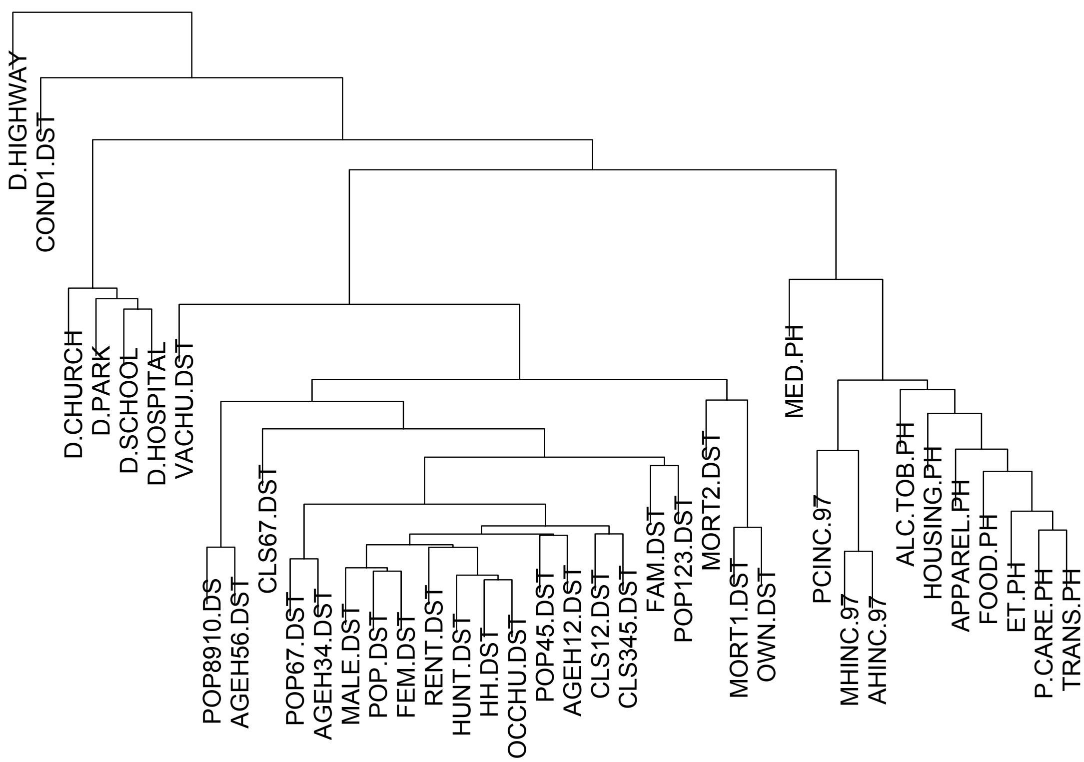


Figure 2. Clusters of features of observed spatial alternatives

From the clustering tree, we divide the features into five clusters. Each cluster includes features with correlated feature values. After checking the distribution of the feature values, we find the feature COND1.DST is almost uniform. It is not a good feature for our analysis. In order to be consistent with Brown and Liu's research (2001), we pick the same features that they used as D.HIGHWAY (distance to highway), FAM.DENSITY (Family density per unit area), P.CARE.PH (personal care expenditure per household) and an extra feature D.HOSPITAL (distance to hospital). These are the key features.

4.3 Model estimation and prediction

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

Using a hot spot model, the priority $p(a_i \in M)$ is estimated with kernel density estimation.

$$p(a_i \in M) = \frac{1}{K}\sum_{k=1}^{K} L\left(\frac{x_i - x_k}{h_1}, \frac{y_i - y_k}{h_2}\right) \tag{8}$$

where x_i, y_i are the coordinates of spatial alternative a_i; K is the total number of observations; L is a function to specify the kernel estimator. We use a Gaussian function here. *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 (1997), $h_i = \frac{4}{(p+2)\cdot K}^{1/(p+2)} \times \sigma_i$ for ith dimension. p is the number of dimensions for density estimation. The estimated model is called the spatial adjusted choice model.

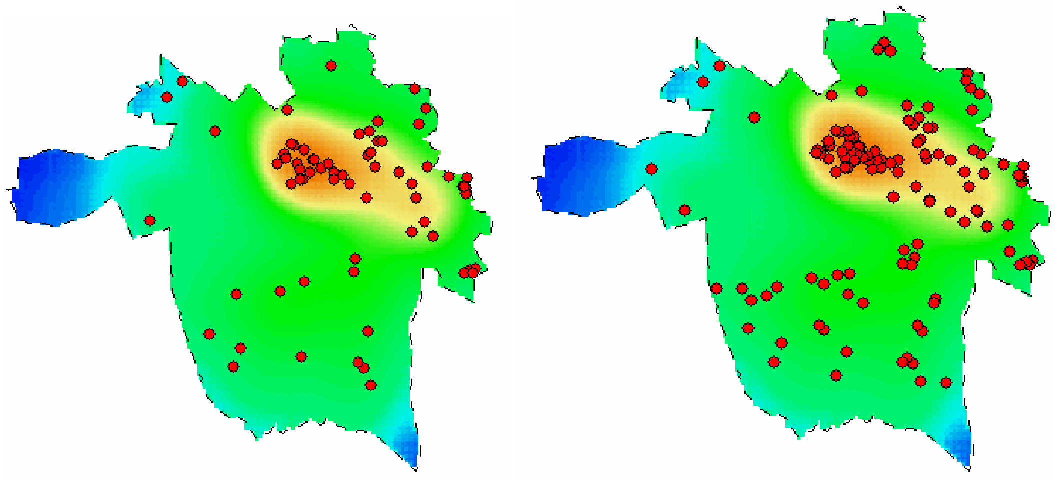
According to previous feature selection processes, we use families per unit area, personal care expenditures per household, distance to highway and distance to hospital as key features. With the density estimation in key feature space, we obtain the prior evaluation probability

$$p(a_i \in M) = \frac{1}{K}\sum_{k=1}^{K} L\left(\frac{x_i^1 - x_k^1}{h_1}, \frac{x_i^2 - x_k^2}{h_2}, \frac{x_i^3 - x_k^3}{h_3}, \frac{x_i^4 - x_k^4}{h_4}\right) \tag{9}$$

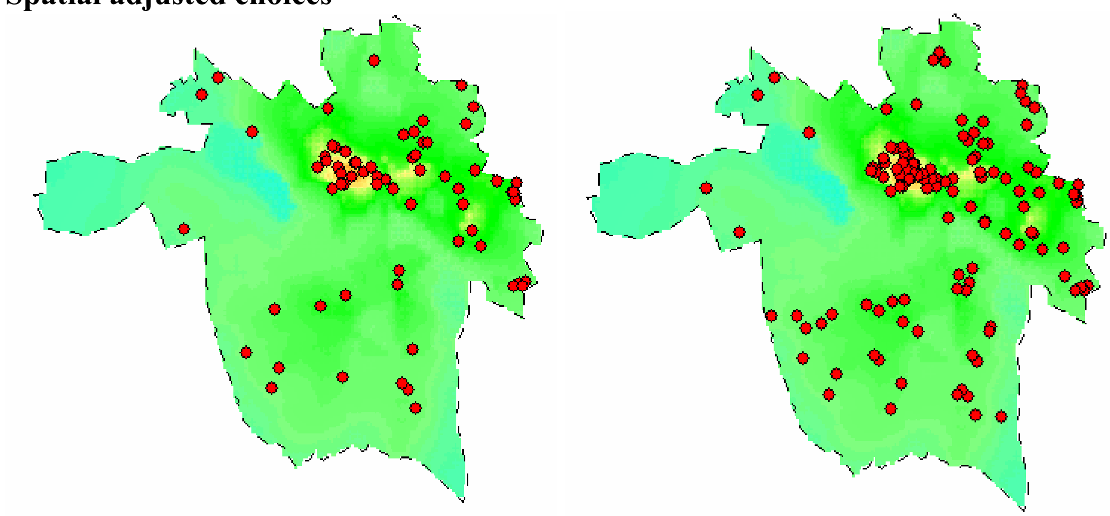
Where, x_i^1, x_i^2, x_i^3 and x_i^4 are the key features of spatial alternative a_i. The estimated model here is called key feature adjusted choice model.

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 multinomial logit models. With the estimation results, we made predictions. Two testing data set 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 spot model and our two spatial site selection models are displayed in Figure 3.

Hot spot mapping



Spatial adjusted choices



Key feature adjusted choices

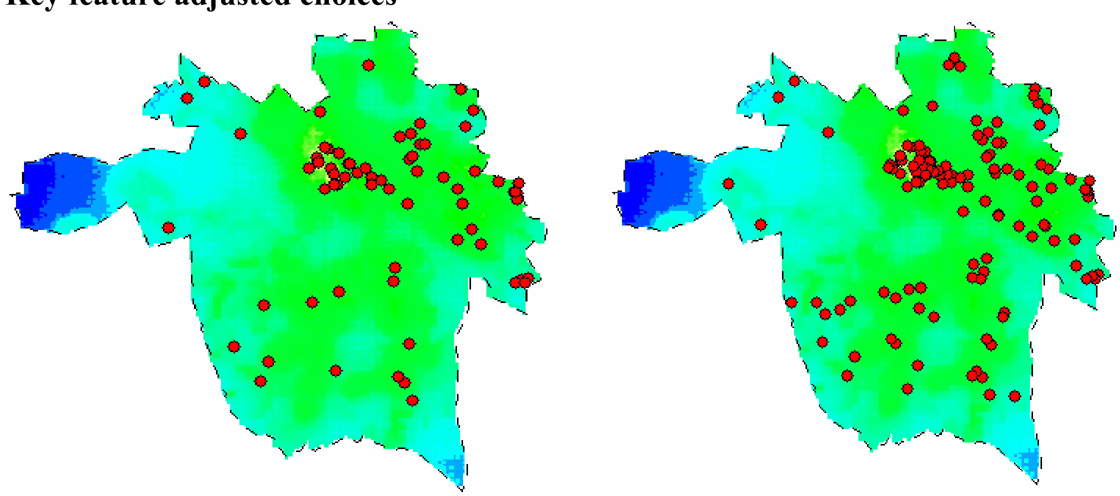


Figure 3. Predictions and crime incidents
between 08/11/97 and 08/17/97 and crime incidents between 08/11/97 and 08/24/97

4.4 Model comparisons

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

Assume that the testing data set contains m incidents that occurred at the locations $\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_m$, respectively. For the incident $\mathbf{s}_i$, let the predicted probability given by the proposed model be $p^p_{\mathbf{s}_i}$ and that given by the comparison model be $p^c_{\mathbf{s}_i}$. The hypothesis test is built around μ which denotes the mean of the difference between the predicted probability given by the proposed model and that given by the comparison model.

Assume that the proposed model will 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 the predicted probability for proposed model will be significantly better than the comparison model. We perform the hypothesis test as

$$\begin{aligned} &H_0\text{: } \mu \le 0, \\ &H_a\text{: } \mu > 0. \end{aligned} \tag{10}$$

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

$$\hat{\mu} = (1/m)\sum_{i=1}^{m}\left(p^p_{\mathbf{s}_i} - p^c_{\mathbf{s}_i}\right) \tag{11}$$

The standard deviation of the difference, $q_{\mathbf{s}_i} = p^p_{\mathbf{s}_i} - p^c_{\mathbf{s}_i}$ is estimated by

$$\hat{\sigma} = (1/(m-1))\sum_{i=1}^{m}\left(q_{\mathbf{s}_i} - \hat{\mu}\right)^2 . \tag{12}$$

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 probability that the null hypothesis will not be rejected.

Table 1. The comparison results

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

	Mean	Std. Dev	z-Statistic	p-Value
Spatial 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. Spatial 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
Spatial 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. Spatial Adjusted	3.977×10^{-4}	9.985×10^{-4}	4.5056	<0.0001

From the comparisons, we find that the two adjusted spatial choice models significantly outperform the hot spot mapping model. The key feature space adjusted spatial choice model also significantly outperforms the spatial adjusted choice model significantly. According to Brown and Liu (2001), the feature space density estimation gives a better prediction of the first evaluated subset of the full choice set of alternatives. Based on the estimation of this subset of the overall choice set in the feature space, we provide a more efficient and accurate prediction method for the prediction of future spatial choices.

5. Conclusion

In this paper, two models for spatial pattern analysis, the spatial adjusted spatial choice model and the key feature adjusted spatial choice model are presented by modifying traditional discrete choice model. Using real crime incidents, we show that the two adjusted spatial choice model can reveal the preferences of unknown criminals in space and give more accurate predictions of spatial patterns of future crimes.

References

Bailey, T. (1994). A review of statistical spatial analysis in geographical information systems. Spatial analysis and GIS. Taylor & Francis Ltd.

Ben-Akiva, M. and Lerman, S. (1985). Discrete choice analysis, theory and application to travel demand. the MIT press.

Bhat C. (2000) Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling. Transportation science, Vol. 34, No. 2, pp. 228-238, May.

Bowman, A. and Azzalini, A. (1997). Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations. Oxford Statistical science series.

Bremer, S. (2000) An exploration of the methods for detecting hot spots and changes in hot spot locations. Master of Science Thesis, University of Virginia.

Brown, D., Liu, H. and Xue, Y. (2001). Mining preferences from spatial-temporal data. Proceedings of first SIAM conference, 2001.

Cameron, A. C., and Trivedi, P. K. (1998). Regression analysis of count data, Cambridge University Press, Cambridge.

Cressie, N. A. C. (1991). Statistics for spatial data. New York: John Wiley and Sons.

Dalton, J. (1999) Bandwidth selection for Kernel Density Estimation of geographic point processes. Master of Science Thesis, University of Virginia.

Fotheringham, S. (1988) Consumer store choice and choice set definition. Marketing Science, Summer, 299-310.

Fotheringham, S., Brunsdon, C. and Charlton, M. (2000). Quantitiative Geography.SAGE Publications Ltd..

Gardner, W., Mulvey, E. P., and Shaw, E. C. (1995) Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial. Psychol. Bull 118:392-405.

Getis, A. (1990). Screening for spatial dependence in regression analysis, Papers of the regional science association, 69.

Haining, R. P. (1990). Spatial data analysis in the social and environmental sciences. Cambridge:Cambridge University Press.

Harries, K. (1999). Mapping crime: principle and practice. NIJ: CMRC. 112-118.

Manski, C. (1981). Structural models of discrete data. Sociogical Methodology. The MIT press.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. Frontiers in Econometrics, 105-142. New York,

McFadden, D. and Train, K. (1978). The goods/leisure tradeoff and disaggregate work trip mode choice models. Transportation research, 12,349-353.

Loiver, M. A. and Webster, R. (1989). A geostatistical basis for spatial weighting in multivariate classification, Mathematical geology, 21, 15.

Osgood, W. (2000). Poisson-based regression analysis of aggregate crime rates. Journal of quantitative criminology. Vol. 16. No. 1.

Ripley, B. D. (1981). Spatial statistics, John Wiley and Sons: New York.

Rust, R. and Donthu, N. (1995). Capturing geographically localized misspecification error in retail store choice models. Journal of Marketing research, Vol. XXXII, 103-110.

Silverman, B. W. (1986). Density estimation for statistics and data analysis. London: Chapman and Hall.

Train, K. (1998). Recreational demand models with taste differences over people. Land economics, 74, 230-239.

Upton, G. J. (1985). Spatial statistics by example Vol. 1: Point pattern and quantitative data, New York: John Wiley and Sons.