Spatial Forecast Methods for Terrorist Events in Urban Environments

Donald Brown, Jason Dalton, and Heidi Hoyle

Department of Systems Engineering, University of Virginia
151 Engineer's Way, Charlottesville, VA 22906
{brown,dalton,hjh7g}@virginia.edu

Abstract. Terrorist events such as suicide bombings are rare yet extremely destructive events. Responses to such events are even rarer, because they require forecasting methods for effective prevention and early detection. While many forecasting methods are available, few are designed for conflict scenarios. This paper builds on previous work in forecasting criminal behavior using spatial choice models. Specifically we describe the fusion of two techniques for modeling the spatial choice of suicide bombers into a unified forecast that combines spatial likelihood modeling of environmental characteristics with logistic regression modeling of demographic features. In addition to describing the approach we also provide motivation for the fusion of the methods and contrast the results obtained with those from the more common kernel density estimation methods that do not account for variation in the event space. We give an example of successful use of this combined method and an evaluation of its performance. We conclude that the fusion method shows improvement over other methods and greater scalability for combining larger numbers of spatial forecasting methods than were previously available.

1 Introduction

Conflict in the modern world has seen a shift from large-scale conventional wars to asymmetrical warfare. Subsequently, recent acts of terrorism and attacks against civilians surfaced due to an imbalance in the weapon systems of warring factions. [1] Before September 11, 2001 attacks against U.S. citizens by foreign terrorists occurred primarily overseas and were typically conducted by young adult males of Middle-Eastern descent. Those trends have changed and the U.S. now realizes increased vulnerability to attacks on its own soil. In response to the events on September 11th, the Department of Homeland Security (DHS) was created with the mission of preventing and responding to future attacks.

The prevention of terrorism is a challenging mission. For example, suicide bombings are one of the most commonly reported acts of terrorism and one that presents enormous difficulties in understanding and prevention. These difficulties arise first because successful suicide bombings include the death of the bomber. Second, the equipment used is simple and easy to acquire so interdiction and tracking of devices is not easy. Third, penetration of the organizations that promote suicide bombing has not succeeded because they recruit based on an understanding of the local populace. Finally, the characteristics of the suicide bombers have adapted to thwart prevention, changing from men to women and, in some cases, children.

H. Chen et al. (Eds.): ISI 2004, LNCS 3073, pp. 426–435, 2004.

Despite these drawbacks to prevention, there are methods that can exploit the patterns in the behavior of suicide bombing organizations. A number of government agencies are looking for these patterns in their databases through the use of spreadsheets and other statistical measures. But there currently exists no method to find the key variables to explain or uncover suicide bombing patterns. The focus of this research is the use of spatial choice analysis to uncover the bombing patterns in past incident locations and to develop an empirical prediction model for future suicide bombings.

2 Background

This section depicts the rising threat of terrorist organizations. Next, it introduces two common techniques for analyzing this type of data; density estimation and spatial data mining. Finally, it gives basic information about the database used in our analysis.

2.1 Terrorist Organizations

To date, only two Middle East terrorist organizations have targeted U.S. interests. The first was the Palestinian Liberation Front (PLF) attack on the Achille Lauro cruise ship where one U.S. citizen died. The second was a series of 1983 attacks where Hezbollah bombed the U.S. Embassy, Marine Barracks, and an annex of the Embassy, all in Beirut. A third attack by a lesser known Israeli terrorist organization, the Harakat ul-Mujahidin (HUM)/Movement of the Holy Warrior, recently abducted and murdered U.S. journalist Daniel Pearl.

Regardless of where these terrorist organizations strike, operating in the U.S. or in Israel, there has been a substantial loss of innocent life. The United States recognized the terrorist threat within its borders and around the world, and subsequently initiated the Global War on Terror.

Agencies are in great need of tools to help them allocate limited resources, proactively engaging and defending against this threat. Because terrorist attacks are rare, we have few data points to make our forecast models. Because they are highly destructive, we need to enact measures that will estimate the threat before an incident occurs so that defensive measures can be taken. There is little response that can be made when a likely target for suicide bombers is found by the bomber before the security agencies.

2.2 Density Estimation

Multivariate density estimation fits a probability density function to empirical data. Techniques for performing density estimation include goodness of fit tests for fitting distributions and nonparametric techniques. Among these nonparametric techniques are clustering methods, e.g., k-means estimation, mixture models, and kernel methods [3]. In a very high dimensional space, it is necessary at times to reduce the dimensionality of the space in order to reduce computation and avoid redundant data elements (multicolinearity). Most current approaches to performing spatial density estimation use only the location of an event, creating a density function in the two coordinate dimensions. In this paper, we demonstrate a method for extracting environmental variables using geographic information systems. These variables more closely model an offender's preferences in target selection, and therefore produce a density of preferred target locations, rather than a density of past event locations.

2.3 Spatial Knowledge Mining

Spatial knowledge mining is a derivative of Knowledge Discovery in Databases (KDD). Spatial knowledge is stored in geographic information systems (GIS), which are databases with additional functionality for performing spatial manipulations and queries from the data. Spatial knowledge mining merges the techniques of data mining and GIS to form high dimensional analysis, the results of which are then projected onto the two dimensional geographical view. This synergy is important because it allows the results of sophisticated multivariate analysis to be presented in an intuitive display for the user.

2.4 Data Sources

The data for this analysis came from several sources. The Israeli regional maps came from the Israeli Central Bureau of Statistics. The suicide bombing event data came from the International Policy Institute for Counter-Terrorism (ICT) at the Interdisciplinary Center in Herzliya, Israel, the largest public international terrorism/terror attack database available on the internet and multiple worldwide news sources. ICT was founded in 1996 in Israel and recently opened an Arlington, Virginia office. Their goal is to "help evaluate the threat that terrorism poses to America and the rest of the world and organize strategic training and orientation activities for officials in the executive, legislative and judicial branches of government." [4] . These data include 517 attacks indexed by date, location, type of attack, organization, and a description of the event. These data were transferred to a GIS and divided into two partitions for model training and testing. Seventy-five percent of the events were used for training and the remaining twenty-five for testing and evaluation. Analysis was conducted using tools developed by the authors for S-Plus, Visual Basic, and ArcGIS.

3 Data Preparation

3.1 Data Extraction

From our raw data, we extracted over 100 candidate features from the spatial environment. From base layers such as the street network, embassy locations, and critical borders we calculated the Euclidian distance from each training event to the nearest feature using GIS. This data was stored in an attribute table for the test set. Due to the regional nature of demographic data, we applied the demographic attributes of the surrounding polygon to each test point. To serve our evaluation and visualization purposes, a reference polygon grid was created in our chosen area of interest. This polygon grid was composed of square cells 50 meters on each edge. The same data extraction steps were performed on the reference grid to derive distance and demographic features. The area of interest grid used in this paper was centered on Jerusalem and contains forty thousand 50 meter cells.

3.2 Feature Reduction – Correlation Structure

In cases where the dimensionality of the data is far larger than its cardinality, a given model cannot give an unbiased estimate of the variance contributed by each feature. This would lead to improper fitting of the model and little robustness. By examining the correlation structure of the data, feature pairs with high correlations can be found and stepwise removed from consideration. By using this method, we can reduce the feature space to a more manageable size, and reduce colinearity of the data set.

3.3 Feature Reduction – Principal Components

Another method of feature reduction is principal component analysis or PCA. PCA seeks to reproject the data onto a new coordinate space such that each successive dimension accounts for a maximal amount of variance of the original process. For a data set of p features, p principal component projections are required to fully and generally account for all of the process variance, but quite often a large portion of the variance can be accounted for by a subset k of feature projections. PCA accomplishes this by projecting each successive feature onto the vector that accounts for the maximum amount of variance remaining. PCA uses the eigenvalue of the correlation matrix to find the projection of data that captures the greatest variance among the data. That projection we call the first principal component, and subsequent orthogonal components are then found and added to the model until we achieve the desired level of cumulative variance.

4 Target Preference Models

4.1 Spatial Preference Model

Most spatial prediction methods use only the past locations to predict future locations (e.g., kriging). In addition, approaches such as density estimation are frequently applied to geography even when assumptions, such as constant variance or homogeneity are violated.

Our approach models spatial preference in a higher dimensional space formed from distances to important geographic or key features. These distances are assumed to have a Gaussian distribution with a mean of zero (i.e, the terrorist would prefer to locate the attack at specific distances from these locations). By measuring the distances of an incident from key features we build an increasingly more precise view of the terrorist preferences. These features are far more descriptive than a geographic coordinate, which is simply an index of the data vector.

We use the index i to indicate a spatial location, and the random variable $\mathbf{D}_{in}$ to indicate the measured distances to key feature i for terrorist attack n. Key features in this problem are features believed to be relevant to decision making by the terrorists, such as government buildings, bus stops, or road junctions.

Next we address the problem of feature selection. The number of possible features is large, if not unbounded. Without knowing the key features we must estimate them using the techniques of principal components and correlation described earlier. These techniques provide either a subset (correlation analysis) or a weighted set (principal component analysis) that provides the basis for predicting the likelihood of an attack.

The likelihood of an attack is given by the density ρ of the location's feature vector in the multidimensional feature space D_i Each event has this vector D with D_i = distance from the event to the nearest key feature *i*. The resulting density function for the distance to feature *i* is given by

$$f(D_{ik}) = \sum_{n=1}^{N} U(D_{ik} - D_{in}) \tag{1}$$

where U(•) is a kernel density operator and k indexes an arbitrary distance and assumes we have discretized over the range of distances.

To evaluate the joint density over the region of interest, assume we have converted the area into a discrete set of 2D points or a grid and let g=1…G be the grid point index. Then the joint density for a single grid point for distances from key features, 1,…,I is

$$f(D_{ig}) = c\prod_{i=1}^{I} f(D_{ig}) \tag{2}$$

c is a constant of proportionality. Substituting from (1) we now have the formulation of the attack likelihood based on spatial preference as

$$\rho(D_g) = c\prod_{i=1}^{I} \frac{1}{N}\sum_{n=1}^{N} U(D_{ig} - D_{in}) \tag{3}$$

Under our assumption of Gaussian uncertainty which implies a Gaussian kernel in (1) the equation in (3) becomes

$$\rho(D_g) = c\prod_{i=1}^{I} \frac{1}{N}\sum_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^{2}}} e^{\frac{-(D_{ig}-D_{in})^2}{2\sigma_i^{2}}} \tag{4}$$

where σ_i is the bandwidth in the Gaussian kernel given by the normal reference density [5].

In this case, this model gives a likelihood for each location, the product of densities for the distances to each feature. When this density is discreetized, the resulting multivariate density can be projected onto the geographic space as a regular grid. This grid is then shaded according to percentiles for visualization. Figure 1 shows 2D and pseudo-3D representations. The points indicate suicide bombing incidents.

4.2 Logistic Regression

Another approach to compute the probability for an attack event is logistic regression. [6] Logistic regression provides a closed form solution to modeling the choice probabilities involved in terrorist site selection and can account for distances to key features as well as categorical variables (e.g., the presence of a holiday).

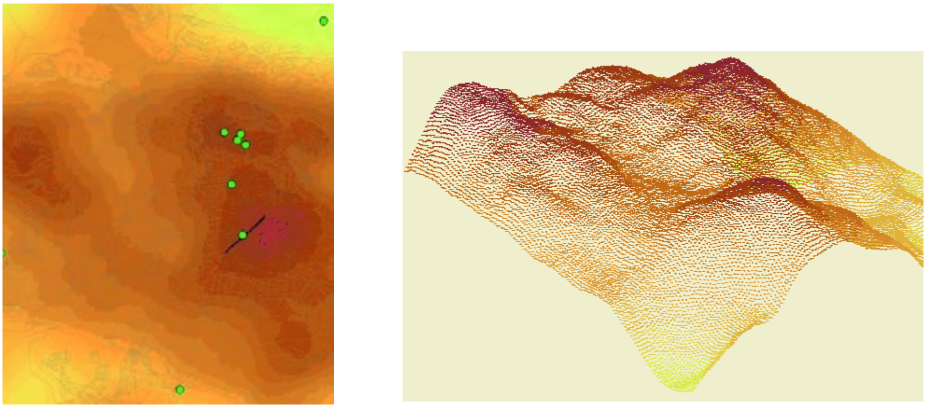

Fig. 1. Spatial Likelihood model

To formulate this model let $\pi_i(x)$ be the probability of a terrorist event at location i given attributes x, where x is a vector of length k.

$$\pi_i(x) = \frac{\exp[B_0 + B_1 x_{i1} + ... + B_k x_{ik}]}{1 + \exp[B_0 + B_1 x_{i1} + ... + B_k x_{ik}]} \tag{5}$$

and 1- $\pi(x)$ is the probability of a non-event:

$$1 - \pi_i(x) = \frac{1}{1 + \exp[B_0 + B_1 x_{i1} + ... + B_k x_{ik}]} \tag{6}$$

so the odds o(x) of an event are:

$$\frac{\pi_i(x)}{1 - \pi_i(x)} = o(x) = \exp\left[B_0 + B_1 x_{i1} + ... + B_k x_{ik}\right] \tag{7}$$

The probability of an event is then compared against a threshold and a classification decision is made.

As noted previously, logistic regression can incorporate variables other than distances and for the prediction of suicide bombings we used demographic attributes. As before, each reference grid cell, i, is given a score between 0 and 1 that represents a likelihood that an event occurs. This grid is then shaded according to similar means as the spatial likelihood model for display purposes. Figures 2 and 3 show the demographic model with its characteristic edge effects rising from the use of aggregate polygons such as census tracts.

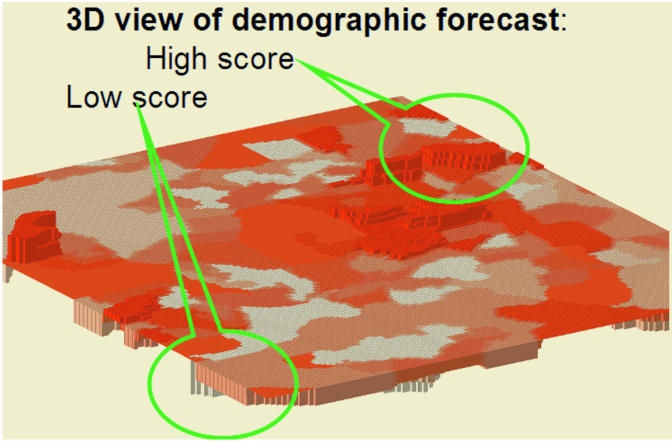

Fig. 2. Logistic regression of demographic model – Perspective

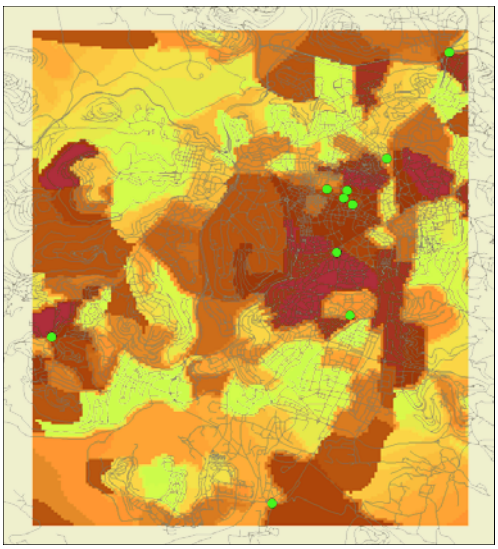

Fig. 3. Logistic regression of demographic model – Bird's Eye

4.3 Fusion of Spatial Models

Because the spatial models are not independent, we cannot fuse the two probabilities as we would two independent sensors. Instead we use the demographic data to augment or detract from the spatial model according to its own score. Because the demographic model is inherently disjoint due to the edges of the aggregation polygons, this merge has the effect of "jig sawing" the spatial model and shifting regions toward higher or lower scores. The areas that appear torn away in the perspective view are regions that have been shifted due to the model fusion. Figure 4 shows this effect.

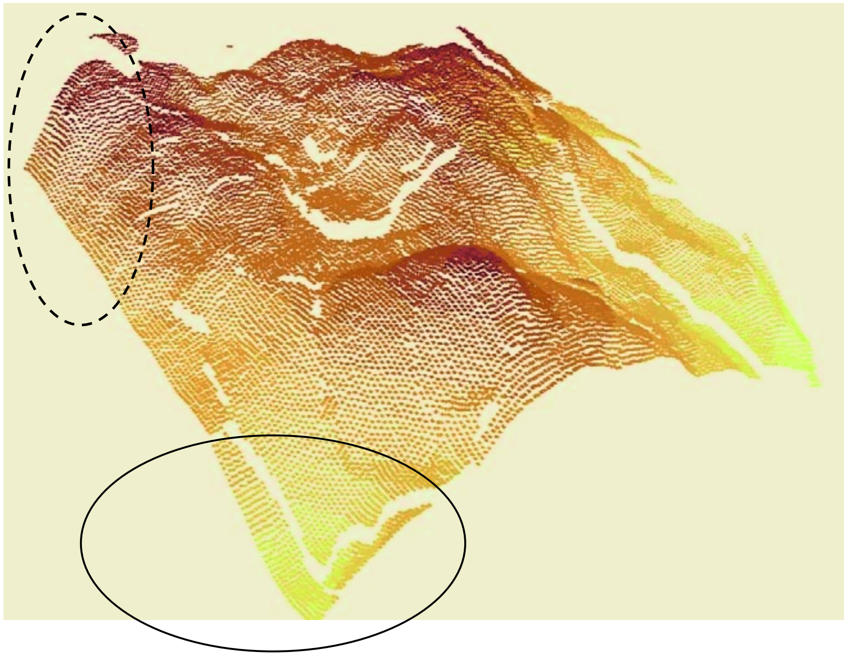

Fig. 4. Aggregate model. Dashed oval shows area of augmented probability. Solid lined oval shows area with reduced probability

5 Evaluation

For the evaluation phase we compare each method using our test set. Spatial forecast evaluation does not have a convenient analog to a simple goodness of fit test as in function estimation. So we use a concept called the percentile score, which computes the percentile given by the forecasted surface for each terrorist event in the test set. The percentile score has values from 0 to 100 and shows the percentage of the estimated distribution that is at or below the value given by the density function at that point. A large value indicates that the predicted surface showed a high likelihood at the actual attack location, while a low value shows that it incorrectly predicted the attack.

Table 1 shows the average percentile [7] scores for the methods described in this paper and also density estimate using Gaussian kernels (KDE) over just the spatial data. Table 1 also shows the sum of the absolute difference between the kernel density estimate with the test set and the estimate provided by each modeling approach. This difference or error shows how well the predictions approximated the surface representation of the actual data.

Table 1. Comparative results

Method	Percentile score	Absolute difference from naïve model (KDE at t_0)
KDE	0.04320	0.4500
Spatial Preference	0.15600	0.3540
Logistic Regression	0.19340	0.3391
Fused	0.27390	0.3069

Table 1 shows that each method in this paper performs better than spatial analysis without key features or demographic information. In addition, the fused approach combines the strengths of the distance measures with the demographic preferences in the logistic regression model to further improve performance.

While the percentile score shows the level of improvement obtained by implementing each model, we also need to formally test the comparisons. We used the Wilcoxon signed rank test as a measure of the statistical validity of our claim of improvement. At a significance level of 0.95, the test rejects the null hypothesis that there is no difference between the model in question and the naïve density model. The alternative hypothesis in all cases is that the density value is higher at all test points. The Wilcoxon tests are required in this data set as we cannot assume independence of the densities. The results generated by the SPLUS statistical package are summarized in table 2 below.

Table 2. Results of Wilcoxon Signed Rank Test

Method	P value	Result
Geographic	0.0002	Alt Hypothesis True
Demographic	0.0001	Alt Hypothesis True
Merged	0.0001	Alt Hypothesis True

6 Conclusion

This paper introduces two new methods for forecasting of spatial point processes. Both methods are individually better forecasters of events than the kernel density estimators commonly in use. By using the non-smooth demographic forecast to augment the spatial forecast, additional improvement is gained over the naïve model. More tests into the ability of the forecast fusion method will determine if there is a significant improvement over either of the two methods individually. While this method was created and implemented for analyzing crimes and terrorist events, there are many applications in resource management and planning that could benefit as well. Currently we are researching the use of the preference modeling techniques used herein to train a behavior engine for agent-based simulation. Further developments will attempt to use the modeling techniques to determine environmental rather than purely spatial bounds for gang activity.

References

1. Cordesman, Anthony H. "The Lessons of Afghansitan: War fighting, Intelligence, and Force Transformation," The Center for Strategic and International Studies, Washington, D.C. 2002.
2. U.S. Department of Justice. FY 2002 Program Performance Report. "Strategic Goal One: Protect America Against the Threat of Terrorism," Downloaded from http://www.usdoj.gov/ag/annualreports/ar2002/sg1finalacctperftpt.htm on October 20, 2003.
3. Parzen, E. "On Estimation of Probability Density Function and Mode," *Annals Math. Statist,* 1962, pp 1065-1076.
4. International Policy Institutes for Counter-Terrorism, "International Terrorism/Terror Attack Database," http://www.ict.org.il/.
5. Venables, W.N.& Ripley, B.D., "Modern Applied Statistics with S," Springer-Verlag New York 2002 p.127.
6. Lattin, James & Carrol, J Douglas, "Analyzing Multivariate Data," Thomson Learning Inc, 2003 Ch 13.
7. Liu, Hua, & Brown, Donald E, "Criminal incident prediction using a point-pattern-based density model," *International Journal of Forecasting*, pp 603-622.