Burglary Crime Analysis Using Logistic Regression

Daniel Antolos[1], Dahai Liu[1], Andrei Ludu[2], and Dennis Vincenzi[3]

[1] Department of Human Factors and Systems Department
[2] Department of Mathematics
[3] Department of Aeronautics, World Wide Campus Embry-Riddle Aeronautical University
Daytona Beach, Florida, 32114
antolosd@my.erau.edu, {dahai.liu,ludua,vincenzd}@erau.edu

Abstract. This study used a logistic regression model to investigate the relationship between several predicting factors and burglary occurrence probability with regard to the epicenter. These factors include day of the week, time of the day, repeated victimization, connectors and barriers. Data was collected from a local police report on 2010 burglary incidents. Results showed the model has various degrees of significance in terms of predicting the occurrence within difference ranges from the epicenter. Follow-up refined multiple comparisons of different sizes were observed to further discover the pattern of prediction strength of these factors. Results are discussed and further research directions were given at the end of the paper.

Keywords: Logistic regression, crime analysis.

1 Introduction

Crime analysis studies certain characteristics of crime and the motives of the persons who committed them developed some form of pattern. Since 1990s, crime analysis has been focused on study the relationship between crime occurrence with human related factors such as personality [17], psychological characteristics[1][4] [7], and factors include peer influence and peer pressure, provocation and anger, boredom and thrill, alcohol and drugs, and money were found to be possible motivators for crimes [5]. Besides the human factors, it has been found that crime activity may have strong correlations to a number of environmental variables that lead to pattern for the occurrence of the crime. For example, how does a criminal, such as a burglar, determine when and where they will target a residence. Nee and Meenaghan [8] found that burglars acted on crime using an almost habitual decision-making process that allow them to navigate quickly and effectively around their world. Pitcher and Johnson [11] showed that a burglar's action is more of a forging and heuristic process when selecting targets. It has been found that burglars consider factors such as the specific period of time for the crime, selection of the target, whether or not the target has been successful in the past and what the terrain is in regards to concealment, ingress and egress. Many factors such as familiarity of the target area [8], distance [14], roadways

S. Yamamoto (Ed.): HIMI/HCII 2013, Part III, LNCS 8018, pp. 549–558, 2013.

and barriers [9] and repeat victimizations on the same target area [15] have been found to play a role in the burglary crime commitment.

Utilizing information attained from historical data, researchers and law enforcement were able to develop models to target the burglar utilizing the characteristics of the crime, such as location, time of day, and day of week [12][13]. Spatial and temporal analysis has been the primary means of determining criminal activity patterns based on the environment and the impact of the behavior upon the environment [6][10]. Burglary behavior and target selection is believed to be predictable by certain characteristics. By identifying these characteristics and analyzing their influences in the decision making process of the criminal one can develop a means to identify potential targets probability that criminal will attempt to target in the future. In turn this will allow law enforcement to focus resources on exploiting key information necessary to reduce or eliminate the criminal threat to a great extent. Although there are many studies that utilize spatial and temporal factors [13][16], but very limited studies represent a combination of factors to predict the probability of crime occurrence with regard to a crime epicenter. This study is intended to address this issue by utilizing a logistic regression model to determine the probability of targeting based on specific conditions of factors

2 Method

2.1 Data Collection

The data were collected from filed police reports from the year 2010 by the local police department. The current crime rate of 7.70 percent resulting in one in thirteen people of having the likelihood of being victimized via burglary, theft, or motor vehicle theft within the city concerning property crime provided enough data to support the study. The data was comprised of the time and date of the offense, approximate start and finish time of the offense, and the street address of the offense.

2.2 Variables

The independent variables for this study were based on prior studies in crime mapping and analysis, including time of day (category data), day of the week (discrete), repeat victimization (the occurrence of the offense at the same residence following the initial offense over the calendar year, category data, 0 or 1), connectors (the amount of access streets, pathways or bridges relative to the targeted residence that allows access to a multi-lane street, discrete variable) and barriers (physical structures that will disrupt or block an individual's egress from a targeted residence, discrete variable).

The dependent variable was the probability of occurrence of crime by the distance from the crime epicenter.

2.3 Model

The logistic regression function denoted as

$$y=f(x)= 11+ e-z \tag{1}$$

With

$$z= \alpha+ \beta1X1+\beta2X2\ldots.\beta kXk \tag{2}$$

With $y = f(x)$ being the dependent variable and $X_i$being the independent variables.

The street addresses from the data were mapped and the coordinates of each incident was determined by developing a grid reference system. By annotating the amount of events per grid, an epicenter can be determined based on the concentration of events within each grid. Incremental diameter circles of 1km were applied from the epicenter to categorize the distance from the established epicenter. This was accomplished by plotting every address in Google Earth and Maps effectively providing the pictorial depiction of the data. The following equations denote the weighted coordinates for the epicenter

$$(i) = \frac{\sum_{j=1}^{n} \sum_{i=1}^{n} i\, v(i,j)}{\sum_{j=1}^{n} \sum_{i=1}^{n} v\,(i,j)} \tag{3}$$

$$(j) = \frac{\sum_{j=1}^{n} \sum_{i=1}^{n} j\, v(i,j)}{\sum_{j=1}^{n} \sum_{i=1}^{n} v\,(i,j)} \tag{4}$$

With i and j being the coordinates and $v(i,j)$ as the number of incidents in the grid as the weights.

Once the addresses were mapped in Google Maps, the amount of barriers and connectors associated with each targeted residence can be determined in combination with Google Earth. A major street system was defined as a multi-lane street with or without a physical medium that can be accessed by turning immediately with the flow of traffic as governed by the local traffic laws. Adjacent neighborhoods were defined by neighborhoods that can be accessed via foot path or connecting residence street systems. Finally it was determined which residences were repeated targets within the given year. This was accomplished from observing the data and annotating which residences were targeted more than once over the course of 2010.

The processed data was then input into SPSS in order to run inferential statistics to determine if any of the variables were significant ($p < 0.05$) prior to conducting a binary logistic regression. A binary logistic regression was conducted utilizing a majority portion ($N = 600$) of the information to determine the probability of observed variables and their relation to a burglary event within certain ranges to the epicenter.

For the validation of the logistic model, a Wald chi-square statistic was utilized to determine the significance of the individual regression coefficients and whether or not they are having some effect upon the regression model. The goodness-of-fit statistic to determine the fit of the logistic model was done utilizing the Hosmer-Lemeshow and supplemental R2 to determine the proportion of the variation in the dependent variable that can be explained by predictors in the model. The resultant probabilities from the model were then revalidated with the actual outcome to determine if the high probabilities are associated with the events in comparison to the incremental one km radii. This was determined by utilizing a measure of association, more specifically a Somer's D statistics.

3 Results

All the events were categorized based on location within the a central Florida city by developing a 1 (km) x 1 (km) grid system to quantify the amount of activity within each cell. Figure 1 illustrates the mapped burglary events with the grid system.). Figure 2 shows that the epicenter was determined to be at 9.10 units along the x-axis and 7.30 units along the y-axis of the grid system.

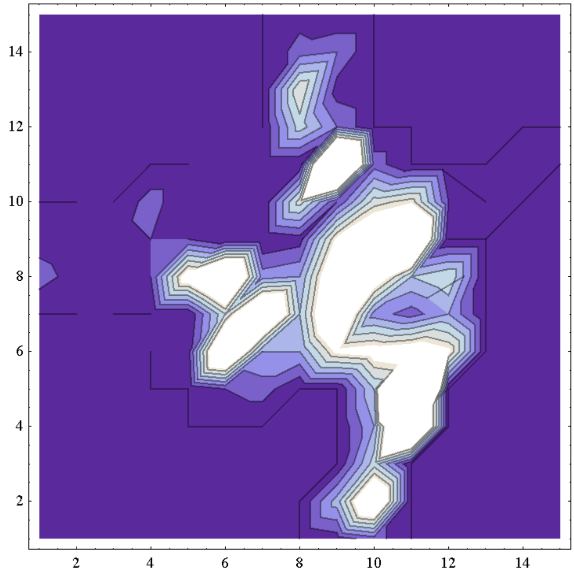

Fig. 1. Graphical representation of grid plotted burglary events depicted utilizing a white hot rendering to show higher concentration in comparison to the low concentration of events as depicted in dark color

A sequence of models were built and tested, starting from one kilometer radius with an increment of one kilometer until the model became insignificant. An overall model evaluation was conducted to determine if the model provided a better fit to an observed data set in comparison to a intercept-only model, or null model which serves as a baseline comparison due to the lack of predictors. It was found the regression models of "one km radii and else", "two km radii and else", and "three km radii and else" were statistically significant. "Four and greater kilometers" radii was not significant. Table 1 illustrates the results for the three significant models.

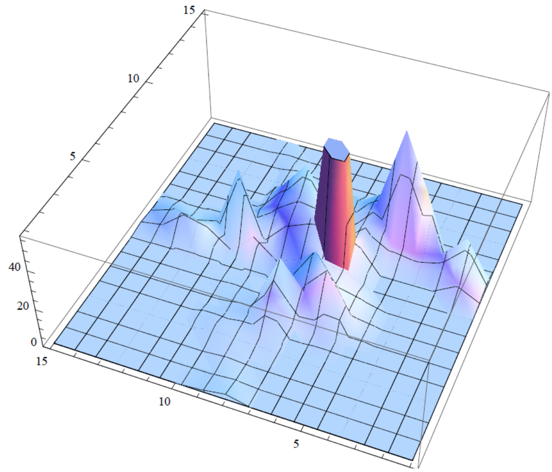

Fig. 2. Graphical representation of epicenter where 9.10 units of x-axis and 7.30 units of y-axis where identified as the epicenter of all the burglary activity in a 3-D rendering

Table 1. Comparison between Logistic Model vs. constant-only model

Radii comparison	< km Sample	> km Sample	α	χ2 testing p value
1 km to > 1 km and greater	23	568	3.207	p < 0.001
< 2 km to > 2 km and greater	132	459	1.246	p < 0.001
< 3 km to > 3 km and greater	283	308	0.085	p < 0.001

Classification tests were conducted of the model to determine the ability of the model accurately predicting the placement of the burglary events as they relate to the radii with which they are associated. The results are depicted in Table 2. Results showed that all the significant models could accurately classify the burglary events in the comparison groups, however overall classification capability of the model diminished as the comparison sizes became close to each other as the radii increased.

Table 2. Classification by radii

Radii Comparison	< km Sample	> km Sample	< km (%)	> km (%)	Overall (%)
< 1 km to > 1 km and greater	23	568	0	100	96.1
< 2 km to > 2 km and greater	132	459	0	100	77.7
< 3 km to > 3 km and greater	283	308	80.6	59.7	69.7

To determine the effects of the individual IVs, a statistical analysis was conducted to determine the significance of the individual regression coefficients. The regression coefficients, Wald statistics, and odds ratios for each of the five predictors were calculated and tested. The results are shown in Table 3. According to the Wald criterion only certain predictors associated per radii comparisons predicted the event of burglary activity. For the three comparison models, connecters were found significant at 1 km with $P[\chi^2(1)>25.286]<0.001$, 2 km $P[\chi^2(1)>3.905]=0.048$, and 3 km $P[\chi^2(1)>9.019]=0.003$; Barriers were found to be significant at 2 km and 3 km comparison models with $P[\chi^2(1)>7.656]=0.006$ and $P[\chi^2(1)>66.692]<0.001$. The rest of the independent variables were not found to be significant for these three models. Table 3 illustrates the significant effect for the individual variables.

Table 3. Significant predictors within the models

Radii	Variables	B	Wald	d.f.	p
1km	Connectors	2.162	25.286	1	0.000
2km	Connectors	0.33	3.905	1	0.048
2km	Barriers	-0.251	7.656	1	0.006
3km	Connectors	-0.404	9.019	1	0.003
3km	Barriers	-0.692	66.692	1	0.000

Table 4. Comparison between Logistic Model vs. constant-only model on 1km band

Radii	Samples within radii	Sample outside radii	Constant α	P value
1 to 2 km	23	109	1.556	$p < 0.001$
2 to 3 km	109	151	0.326	$p = 0.003$
3 to 4 km	151	82	-0.611	$p < 0.001$
4 to 5 km	82	136	0.506	No significance
5 to 6 km	136	45	-1.106	$p < 0.001$
6 to 7 km	45	9	-1.609	$p < 0.001$
7 to 8 km	9	32	1.003	$p < 0.001$
8 to 9 km	32	4	-2.015	P=0.007

Furthermore, in order to test the sensitivity of small scale of 1 km radii on the prediction of the crime occurrence, eight different models on scale of 1km band (1-2km, 2-3km, etc.) was constructed. It was found that even the small scales, the predictors, as a set; reliably distinguish between the different radii as they increase by one kilometer from the epicenter to the nine kilometer radius. Results of the nine different models are illustrated in Table 4.

Classification varied by kilometer comparison being relatively strong from one to four kilometers and five to nine kilometers with results ranging from 74% to 83%, except for the two to three kilometer comparison resulting in 53%. These classifications depict the ability of the models to correctly classify the events per the kilometer radii based on the five independent variables.

According to the Wald criterion, only certain predictors associated per radii comparisons predicted the event of burglary activity. Connector and barriers were significant predictors from one to three kilometers from the epicenter and from five to six kilometers. Repeat victimization was only significant within one to two kilometers and day of week from five to six kilometers.

4 Discussions

This study utilized logistic regression models to predict the probability of burglary activities with respect to the event density epicenter. The predictors of time of day, day of the week, connectors, barriers, and repeat victimization were selected. Results showed that certain variables such as connectors and barriers had significant effects when observed in certain radii size comparisons; however there was variation in the predictors of time of day, day of week, and repeat victimization.

For the models for the radii size comparison of one kilometer to the remaining, two kilometers to the remaining, and the three kilometers to the remaining data population were found to be significant, the bigger radii were insignificant for the prediction. As shown with additional analysis with individual factors and smaller scale model comparison, it was found that the results have a large degree of variability, which produce a wide range of probabilities for the occurrence prediction. A number of factors could be attributed to this variability.

One, the burglaries in the local population area has a large degree of randomness, due to the nature of tourism for the city and weather conditions. As a result, the burglary events in the different radii comparison groups the variation between the data continued to become too great to determine any significance within the time of day and day of the week predictors.

Secondly, Nee and Meenaghan (2006) found that burglars had multiple reasons for committing a crime; money, thrill, and to support a drug habit could be the majority of reasons. However, they found that the individual would search out an area when the need to commit the crime had to satisfy one of the aforementioned reasons, which can be best described as stochastic. It is possible that there might be no discernible pattern or set of circumstances that would allow the burglar to target a specific residence. In our study, the data was compiled from all burglary activities within the lcoal area, which implies that the decisions of multiple burglars could involve with other factors, causing the complexity for predicting the occurrence accurately.

In this study, the level of detail associated with repeat victimization was not explored precisely. Based on the layout of the local area, the repeat victimization was often not the action of the same offender but of different burglars to generate an effect of repeat victimization. The addition of multiple offenders within a given area without prior knowledge and history of the targeted location cause the deviation from the pattern generation for this factor.

Furthermore, there is a massive single lane road system that interconnects neighborhoods of local area, essentially eliminating the need for the multi-lane road system for egress from a targeted location. The size of the neighborhoods in local area also affords additional concealment by allowing a burglar to access different areas effectively expanding the area searched by law enforcement following a crime and reducing detection. The unique features of connectors and barriers for the local area might lead to its significance for the model.

Hammond and Young [3] showed that crime will effectively reduce as it expands from the identified epicenter, consistent with the findings from our study. However one has to keep in mind that the identified epicenter was calculated based on all empirical data. Based on observation the majority of the crime took place in a high demographic area that is characteristic of low economical and financial income which results in 15% of the population falling below the poverty level. This is greater than the national average as reported by the 2009 U.S. Census Bureau report. By incorporating all the data from local police department, it eliminated significant effects that would normally be demonstrated by specific types of demographic, economics, and financial influences, as Haddad and Moghadam [2] showed that isolating a specific demographic and economics associated with a particular area would yield stronger results when determining the contributors associated with burglary. The mixture of data from different areas could lead to some diminishing patterns for the crime.

5 Conclusions

The study used logistic regression to investigate the factors associated with burglary activity for a local city. It was found that only repeat victimization, connectors, and barriers were significant in determining a burglary probability at certain radii ranges. Further analysis and discussion suggests that the size of the observation area of city and other factors must be refined to strengthen the predictive model. It is believed that the local police department would benefit from this study by utilizing the model within a more constrained area observing the same factors as described in this study. Applying the procedures and model as discussed in this study within a one kilometer area with a 100 meter radii incremental comparison, will provide the resolution required for predictive analysis.

The limitation of this study also pertains to the method itself. Logistic regression model is limited to identifying the probability of burglary activity. It only provides information as it relates to the likelihood of burglary activity within certain radii distances of the cluster's epicenter. This model will not identify specific locations, specifically residences, but will encompass the roads associated with an identified area in turn contributing to the police department patrol plan to deter or capture burglary suspects.

Given the conclusion that smaller geographic areas are required for stronger time and space analysis, connectors and barriers potentially no longer will be influential given the observation the burglar will remain in a neighborhood they too utilize for residence. On the contrary, connectors and barriers can be further refined to determine the amount of association the targeted residence has with the actual residence of the described offender. Overall greater detail needs to be incorporated into a study that will utilize the predictors of time of day, day of the week, connectors, barriers, and repeat victimization. As demonstrated by this study encompassing a wide range of data over a large area with no refinement will yield insignificant results. This study can further be refined to observe smaller clusters of activity; this will allow law enforcement professionals to focus patrol efforts within a one kilometer residential

In reference to utilizing this study for other modes such as other law enforcement activities or IED pattern analysis and prediction the same approach can still be utilized when specific targeted areas of interest are identified based on activity and proximity to one another. The reduction of the observed area along with detailed analysis of the events within that given area will increase the significance of the predictors as well as develop an accurate prediction model.

References

1. Gottfredson, M., Hirschi, T.: A General Theory of Crime. Stanford University Press, Stanford (1990)
2. Haddad, G.R.K., Moghadam, H.M.: The socioeconomic and demographic determinants of crime in Iran (a regional panel study). European Journal of Law and Economics 32(1), 99–114 (2011)
3. Hammond, L., Youngs, D.: Decay functions and criminal spatial processes: Geographical offender profiling of volume crime. Journal of Investigative Psychology and Offender Profiling 8(1), 90–102 (2011)
4. Horney, J.: An alternate psychology of criminal behavior. Criminology 44, 1–16 (2006)
5. Kai, Y.T., Heng, M.A., Bullock, L.M.: What provokes young people to get into trouble: Singapore stories. Preventing School Failure 51(2), 13–17 (2007)
6. Jing, W., Tao, Z.: Analysis of decision tree classification algorithm based on attribute reduction and application in criminal behavior. In: Pro. of 2011 3rd Conference on Computer Research and Development (ICCRD), Shanghai, China, March 11-13, pp. 27–30 (2011)
7. Mischel, W., Shoda, Y.: A cognitive-affect system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. Psychological Review 102(2), 246–268 (1995)
8. Nee, C., Meenaghan, A.: Expert decision making in burglars. Brit. J. Criminol. 46, 935–949 (2006)
9. Peeters, M.P.: The influence of physical barriers on the journey-to-crime of offenders. Netherlands Institute for the Study of Crime and Law Enforcement, The Netherlands (2007)

10. Pillai, G., Kumar, R.: Simulation of human criminal behavior using clustering algorithm. In: Proc. of International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007), vol. 4, pp. 105–109 (2007)
11. Pitcher, A., Johnson, S.: Exploring theories of victimization using a mathematical model of burglary. Journal of Research in Crime and Delinquency 48(1), 83–109 (2011)
12. Ratcliffe, J.H.: The hotspot matrix: a framework for the spatio-temporal targeting of crime reduction. Police Practice and Research 5(1), 5–23 (2004)
13. Ratcliffe, J.H., McCullagh, M.J.: The perception of crime hot spots: a spatial study in Nottingham. In: LaVigne, N. (ed.) PERF/Crime Mapping Research Center, U.K., Washington, DC, ch. 6 (1998)
14. Rengart, G.F.: The journey to crime. In: Bruinsma, G., Elffers, H., Willem, J., de Keijser (eds.) Punishment, Places, and Perpetrators: Developments in Criminology and Criminal Justice Research, pp. 169–181. Willan Publishing, UffculmeCullompton (2004)
15. Sagovsky, A., Johnson, S.D.: When does repeat burglary victimization occur? The Australian and New Zealand Journal of Criminology 40(1), 1–26 (2007)
16. Short, M.B., D'Orsogna, M.R., Brantingham, P.J., Tita, G.E.: Measuring and modeling repeat and near-repeat burglary effects. Journal of Quantitative Criminology 25(3), 325–339 (2009)
17. Sterzer, P.: Born to be criminal? What to make of early biological risk factors for criminal behavior. The American Journal of Psychiatry 167(1), 56–60 (2010)