Analysis of Criminal Behaviour Using a Logistic Regression Model

Maria Jofre
Department of Industrial Engineering, University of Chile
mjofre@ing.uchile.cl

Introduction
Understanding the crime phenomenon is fundamental for crime fighting institutions. Therefore, statistical evidence has acquired a critical importance when analysing criminal behaviour. The distribution of thefts varies over space, suggesting that urban characteristics are key variables when a crime occurs. This study provides analytical evidence of the impact of urban attributes in the expected level of crime.

Method
Data Collection
Two databases are used in this study: (i) police crime reports of thefts, and (ii) urban attributes related to commerce, tourism, culture, transport, government and religion, among others.

Data Processing
The studied region is divided in smaller areas defined by Voronoi polygons centred on street intersections.

Logistic Regression Model
To achieve the pursued objective, a multivariate logistic regression model is proposed. The idea is to determine the relationship between a dependent variable (Y) and one or more independent explanatory variables (X). In this case, the dependent variable is the level of crime and the independent variables are the urban attributes distributed along the studied area.

The model determines a coefficient ($\beta$) for each attribute ($i$), that defines the contribution to the predicted probability of high level of crime ($P(Y=1)$), which is calculated as follow:

$$P(Y = 1) = \frac{1}{e^{-\beta \cdot X(i)} + 1}$$

For this particular data mining technique, a Maximum Likelihood method is used to estimate the coefficients of the model. The idea of this method is to find the values of $\beta$, that maximise the probability of the real data being generated by the model.

Subsequently, the estimated coefficients are used to calculate the probability of theft occurrence in each Voronoi polygon.

Once the probability of occurrence is defined, each polygon is classified in deciles depending on how likely is to present a high level of crime.

Results
The application of the proposed model determines the relationship between 18 considered urban attributes and the expected level of crime.

The results show that 6 urban characteristics are the most important to explain the phenomenon. All of these attributes are directly proportional to the level of crime. Therefore, if these variables are present in a particular area, then a higher amount of thefts is expected in that region.

Conclusion
The presence of chemists, pedestrian walkways and shopping centres define a greater likelihood of a theft.

Given that the police resources are limited, it is recommended police surveillance in critical areas determined by the model.

Also, we recommend using the results of this study for urban planning, taking in consideration the most dangerous combination of urban characteristics.

As a future work we expect to model crime as a spatiotemporal phenomenon using more complex statistical techniques, such as Gaussian Process. This model can also quantify the uncertainty in crime prediction.