Crime Data Mining: Combining Socio-economic and Spatial Analysis

Ricardo Ruíz[1], Christopher R. Stephens[2], and Santiago Roel Rodríguez[3]

[1] Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas,
UNAM, DF,
Mexico

[2] Instituto de Ciencias Nucleares, C3 – Centro de Ciencias de la Complejidad,
UNAM, DF,
Mexico

[3] RRS y Asociados, SC,
Mexico

d_ruiz_r@uxmcc2.iimas.unam.mx, stephens@nucleares.unam.mx,
sroel@hellohelp.org

Abstract. Public security is an important issue for a society. With the massive increase in electronic data availability over the last few years, characterising and predicting crime has become a task that can be approached using different data mining techniques. However, previous studies have concentrated on the spatial patterning of crimes without investigating potential predictors or causes. In this paper, we use a range of data mining techniques to analyse criminality using a data base of crimes from the municipality of General de Escobedo in Nuevo León, México. We show that different types of crime - domestic violence, residential robberies and business robberies - have quite different profiles, both from the point of view of the characteristics of the robberies themselves and from the underlying socio-demographic and socio-economic factors that influence them. We create predictive models for these three crime types and discuss how the results can be used to predict and reduce crime risk.

Keywords: Data mining, Bayesian analysis, crime, spatial data mining, crime types

1 Introduction

Crime and security are issues that are almost always at the forefront of the publics mind. There are a wide variety of crimes that can afflict a society or a certain population: from business robbery, home robbery and domestic violence to homicides and kidnappings. There are many questions that can be posed that are relevant to understanding both crime patterns and also the risk factors associated with different crimes. Are all populations equally at risk for

a given crime? What are the underlying socio-demographic and socio-economic factors associated with a given crime? In this big data age, recently, data mining techniques have been brought to bear on this type of problem [1–5]. Many of these studies have focused on the spatio-temporal patterns of crime using unsupervised learning. Such analyses, although providing useful intelligence, do not make potentially causal links to underlying socio-economic and socio-demographic variables as potential risk factors that characterize or profile a particular type of crime. There is also the related question of to what extent one type of crime differs from another in terms of its predictive profile. For instance, are the predictive drivers of all types of robbery the same or do they differ between one robbery type and another? In this paper, we consider those three crime types with the greatest incidence in the municipality of General de Escobedo in Nuevo Leon, México. The three crimes types were: domestic violence, business robbery, and burglary. Criminal data were obtained from the reports of the police officers who attended the crimes. These reports contain valuable information, such as the hour, day, week, month etc. of the crime. However, they do not contain a description of the type of population which is affected by the crimes. This information was obtained from the AGEBs (Basic Geostatistics Area, in English) provided by INEGI [10]. We performed several different types of analysis, ranging from basic exploratory analysis, using simple statistics, to a more sophisticated classification model using a Naive Bayesian classifier. The classifier was used to determine risk profiles for different crime types and perform a spatial risk analysis at the level of AGEBs using, for example, heat maps as a visualisation tool.

1.1 Data Characteristics

The data which we analysed comes from reports made by police officers of the municipality of General de Escobedo in 2012. There are 17 distinct crime types in the data with the highest incidence being associated with domestic violence, business robbery and house robbery. In Table 1 we show the frequencies of the different crime categories.

The original data displayed many inconsistencies and errors, such as typographical errors, domain errors in an attribute and a variation of the field format in the months. As with most data mining projects a substantial amount of time was spent cleaning the data.

1.2 Preliminary Analysis of the Data

We will first illustrate the type of exploratory analysis that is possible with this type of data restricting attention to the case of BR (Business Robbery). BR was chosen due to the greater precision with which it could be located geographically and due to the fact that the associated data was of higher quality - less data errors or missing fields. Similar analysis was performed for other crime types but will not be presented here due to space restrictions.

Table 1. Frequencies of the distinct crime categories

Abbreviation	Number	Type of crime
HOMICIDE	21	Homicide
AR	27	Attempted Robbery
W	27	Breaking and entering
ST	52	Simple Theft
JD	70	Judicial Dictum
CPD	72	Car Pieces Robbery
DP	79	Damage to Property
A	100	Another
I	106	Injuries
G	134	Gangs
FT	160	Failed Theft
VTR	213	Vehicle with Theft Report
VR	237	Vehicle Robbery
TP	290	Theft from person
HR	419	House robbery
BR	675	Business Robbery
DV	1047	Domestic Violence

Each BR event was associated with a small set of descriptor variables. From these variables initial preliminary analysis could be carried out. The BR characteristics which we worked with are:

X_1: Police shift - morning, evening, night, and early morning.
X_2: Time at which crime was reported
X_3: Colonia (Neighborhood)
X_4: Codigo postal (Zip code)
C : type of business robbed

Some representative results can be seen in Figures 1 and 2. In Figure 1 we see an increased incidence for BR at weekends and in Figure 2 an increased incidence at night time. With respect to other features, such as month of the year, day of the month etc. no relevant trends were found in the preliminary analysis. However, it was seen that different types of business exhibited different profiles in terms of the above variables. For instance, convenience stores were much more likely to be robbed at night when compared to other business types.

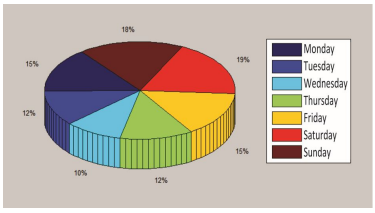

Fig. 1. Percentages of BR by day of the week

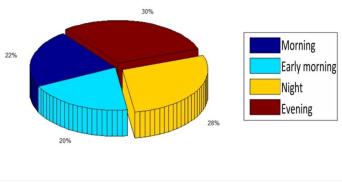

Fig. 2. Percentages of BR by time of day

2 Feature Selection and Model Construction for BR

In order to determine which features are correlated with and therefore potentially predictive of a given crime we must look for the characteristics X_i which are correlated with the crime type considered as a class C. In order to determine these features that are relevant for classifying and predicting a given type of crime we use the statistical diagnostic in equation (1).

$$\varepsilon\left(C|X_i\right) = \frac{N_x\left[P\left(C|x\right) - P\left(C\right)\right]}{\left[N_x P\left(C\right)\left(1 - P\left(C\right)\right)\right]^{1/2}}, \tag{1}$$

where $P(C|x) = \frac{Nc(x)}{Nx}$ and $P(C) = \frac{N(c)}{Nt}$ and:

N_x=number of times that x appears,
$P\left(C\right)$=is the probability that my class has,
$N_c\left(x\right)$=is the number of times x appears in my class,
$P\left(C|x\right)$=the probability of belonging to the class given that we have the characteristic x,
$N\left(C\right)$=number of times that it my class appears,
Nt=total number of records that we have.

ε is a binomial test that determines the statistical significance of the observed distribution of co-occurrences of the class C and feature X_i relative to the null hypothesis that they are uncorrelated. The associated distribution is binomial but may be approximated in most circumstances by a normal distribution. In this case, $|\varepsilon\left(C|X_i\right)| > 1.96$ corresponds to the 95% confidence interval that the observed co-occurrence of C and X_i would not have occurred by chance. Hence, we may determine which features are correlated with or predictive of a given crime type.

Having determined those features which are correlated with the class C we may construct a classifier (score function)

$$S\left(\mathbf{X}\right) = \ln\frac{P\left(C|\mathbf{X}\right)}{P\left(\bar{C}|\mathbf{X}\right)}, \tag{2}$$

where $\bar{C}$ is the complement of the class C and $\mathbf{X}$ is the vector of relevant features. As $P(C|\mathbf{X})$ is potentially a high-dimensional joint probability it cannot be estimated directly. We can proceed however using Bayes rule and adopting the Naive Bayes approximation. Using Bayes rule

$$P(C|\mathbf{X}) = \frac{P(\mathbf{X}|C)P(C)}{P(\mathbf{X})}, \tag{3}$$

we wish to determine the likelihood function $P(\mathbf{X}|C)$. As mentioned, due to the high-dimensional nature of the feature space these likelihoods cannot be determined directly. In the Naive Bayes approximation we assume that the features X_i are not correlated therefore

$$P\left(\mathbf{X}|C\right) = \prod_i P(X_i|C) \qquad P\left(\mathbf{X}|\bar{C}\right) = \prod_i P(X_i|\bar{C}). \tag{4}$$

We thus obtain for our classifier

$$S(X) = \ln \frac{\prod P(X_i|C)}{\prod P\left(X_i|\bar{C}\right)} + \ln \frac{P(C)}{P(\bar{C})}, \tag{5}$$

where $\ln \frac{P(C)}{P(\bar{C})}$ is a constant independent of the features X_i.

A high positive score value indicates a higher probability of belonging to the class C, while a high negative score indicates a low probability of belonging to the class. In the case at hand, in order to use this formalism it is necessary to be able to identify events in $\bar{C}$. However, the provided data base consisted only of crime events not "non-crime" events! There are several ways to overcome this. In the case of BR for instance, if one had access to a data base with all businesses in the municipality one could determine which businesses had not been robbed and these businesses would form $\bar{C}$. Another way to proceed, whch we will adopt here, is to consider C to be a subset of crimes within a wider set. For instance, BR could be the class C and $\bar{C}$ the class of all robberies that were not BR. In this way we will characterize and profile crimes one relative to another rather than in absolute terms. Thus, for instance, we will determine what are the particular predictive drivers of BR relative to other types of robbery.

3 Socio-demographic Variables

By having just few variables that can characterize the different crimes, we looked for other sources of information that would have a better characterization of the offences. The source of this new information is the AGEBs (Basic Geostatistics Area, in English) with which you can get 188 variables that describe the population in a particular area.

For the municipality of General de Escobedo INEGI has a division of 121 AGEBs, not all the territory of the municipality is divided, there are areas, mostly in the periphery, where there are irregular settlements that do not have assigned a AGEB, causing that some socio-demographic data cannot be assigned for some reported crimes.

The variables are grouped into population, migration, indigenous population, disability, educational characteristics, economic characteristics, health, marital status, religion and housing; the vast majority of these variables are accounted with regard to the number of people with those characteristics.

4 Models

The first we did was to geo-code each crime and determine in which AGEB they belong to. From 121 AGEBs in total, we could locate at least one crime in 67 AGEBS, for the remaining AGEBs, there are not records of any crime, which does not mean that the crime was not committed, but is not possible to establish the offence-AGEBs relationship. The possible factors for the preceding situations are:

a The AGEB is an unpopulated area.
b Lack of information regarding that area.
c Irregular settlement.
d Inconsistent information to geographically locate the crime.
e The crime was not reported to the corresponding authorities.

The previous data are some aspects why you cannot set a relationship between the crime and AGEB.

From the 67 AGEBs with at least one offence a coarse-grain is done due to the fact that each variable has a unique value by which a count could not be done by the value in the variables of the AGEB.

Table 2. COARSE-GRAIN.

AGEB	POBTOT	POBMAS	STANDARDIZATION	COARSE-GRAIN
0043	4435	2165	0.488162345	1
0062	151	70	0.463576159	1
0151	3857	1975	0.512056002	8
0166	2348	1155	0.491908007	1
0185	1057	535	0.50614948	5
0202	5293	2621	0.495182316	2
0221	4279	2184	0.510399626	7

Table 2 shows how coarse-grain takes place: first it must be normalized each of the variables, by which POBTOT which is the total of population should be divided between POBMAS which is the total male population; that is how we got the column STANDARDIZATION. This is repeated for the 67 AGEBs and for the 188 variables, then each column corresponding to each variable STANDARDIZATION should be ordered from the least to the greatest, then proceeds to divide the column of STANDARDIZATION into equal sections. In this case as they are 67 AGEBs it was determined to divide them into 8 groups, with which the order creates a new coarse-grain column and assigns the value 1 to the first 8 records, the value 2 to the following 8 records, and so on to assign the value 8 in the last records. Finally a join is done with the records of crimes.

4.1 Domestic Violence Model

To create this model we applied the equation 1 and is taken as "domestic violence" class for the type of offence.

The variable male population from 15 years and over with primary school (last level of studies completed) with the value of 8 has the highest epsilon, 8 value indicates that it is in the coarse-grain that contains more people of this type. What can we infer from the information above?

Table 3 which contains the 10 records with the highest epsilon, we can see clearly the trend of the domestic violence crime with respect to the level of education; most of the population have either primary completed or incomplete,

Table 3. Variables with higher epsilon for family violence .

Variable	Value	Epsilon
Population 15 years old and over with unfinished secondary school.	6	9.0692
Private inhabited house with car property.	1	9.0940
Private inhabited house with fixed telephone line.	1	9.1247
Male population from 0 to 2 years.	6	9.2146
Population from 12 years old and over.	3	9.3088
Male population from 15 years old and over with primary incomplete.	7	9.3385
Female population aged 15 years old and over with primary school finished.	7	9.3845
Female illiterate population from 15 years old or more.	7	9.6971
Male illiterate population from 15 years old or more.	7	9.9840
Male population from 15 years old and over with primary school completed.	8	10.0506

there are some even illiterate. The previous information leads to an economic adverse situation by not having sufficient studies to get a well-paid job.

This is confirmed by the variables private inhabited house with car property and private inhabited house with fixed telephone line, both with a value of 1. Which indicates the existence of population with low purchasing power so that it is below the level of houses with a car or without telephone line, which also indicates a high degree of marginalization.

Table 4. Variables with higher epsilon for family violence .

Variable	Value	Epsilon
Private inhabited houses with one bedroom.	6	7.7072
Population with no health service.	6	78.016
Population from 12 to 14 years old who do not attend school.	6	7.8171
Male population born in another entity.	8	7.9968
Private inhabited houses with two bedrooms or more.	2	8.0916
Private inhabited houses with a computer.	1	8.4723
Private inhabited houses with piped water.	2	8.4739
Occupants per room average in private inhabited houses.	6	8.6375
Illiterate population from 15 years old or more.	7	8.8522
Population 0-2 years old.	7	9.0056

Table 4 shows other relevant variables for domestic violence (these are not the variables that continue in table 3) as you can see, the trend of the economic level is maintained. The variable male population born in another entity with a value of 8 indicates that the majority of the population has emigrated to the municipality, which means they have to rent a room or, in a better case, a social interest housing, there is also the possibility of arriving with a family member, which looks reflected in the variables average of occupants per room , in private occupied houses, private inhabited houses with a bedroom, both with a value of 6, which shows a tendency to have many people living in a small space, while the variable private inhabited houses with two bedrooms and more with value of 2 shows that the majority of dwellings where there is high incidence of domestic violence are houses with no more than two bedrooms. Having more people living

in reduced inhabiting spaces, in precarious economic situation and low education level leads to high rates of domestic violence.

In Figure 3 there is a heat map regarding the incidence of domestic violence in each AGEB (red color indicates a higher incidence, white color means little or no presence) as we can see there is a clear trend to the furthest areas from the municipality which are belonging to the periphery of Monterrey city, adjacent to uninhabited areas, some are irregular settlements. The mentioned areas coincide with the variables of the previous analysis.

In the table 5 we have 10 variables with the more negative epsilon, as it is visible, it can be considered the against part of table 3. For example, the variable private inhabited houses with fixed telephone line with the value of 8 and epsilon - 8.95, it is the opposite to the variable with the same name in the table 3 but with value 1 and epsilon 9.124. In the table 5, we can see that variables which involve education are those that indicate a higher education level than those with only basic education. All these variables have a value of 8 or 7 and belong to the coarse-grain that brings a greater number of people who have the post-basic education. On the other hand, the variables of education indicate an incomplete or complete basic education with a value of 1 which belong to the coarse-grain with the least number of people of this type. The private houses inhabited with a washing machine with a value of 8 indicates a socio-economic level better than those described in the variables of table 3.

This indicates that the incidence of domestic violence is far less than expected as the level of education rises, the same applies to the economic level.

Table 5. Variables with lower epsilon for family violence.

Variable	Value	Epsilon
Male population from 18 years old and over with post-basic education.	8	-9.3353
Private inhabited houses with fixed telephone line.	8	-8.9527
People from 18 to 24 years old attending school.	7	-8.8008
Population from 18 years old and over with post-basic education.	8	-8.7268
Female population from 18 years old and over with post-basic education	8	-8.7268
Female population from 15 years old and over with secondary school incomplete.	1	-8.6027
Population without religion.	1	-8.4974
Private inhabited houses with radio.	8	-8.4072
Private inhabited houses with a washing machine.	8	-8.4072
Female population from15 years old and over with secondary school complete.	1	-7.9293

4.2 Business Robbery Model

Now, we will analyze the model for the crime business robbery and determine what characterizes this offence from others.

As seen in table 1 657 businesses robberies are registered, in table 6 you can see the first 10 variables with higher epsilon, there are four variables 10 regarding the location. This allows us to infer that business robbery is characterized by the geographic area of incidence. The rest of the variables tells us where to open

a new business more than characterize the business robbery, in other words, the variables as private inhabited houses with washing machine, private inhabited houses with fixed telephone line with a value of 8, all these data indicates the economic level of the area where a business is established. So that, in order to exist a business robbery there must exists a business there. So, where a business must be set?, In a low or in a high socio-economic level area?, It is clear that in an area where the purchasing power of people allows them to purchase products or services offered by a business. Therefore, economic variables, which characterize a business robbery, determine where it is it more feasible to find a business or where people can open one.

Table 6. Variables with higher epsilon for Business Robbery .

Variable	Value	Epsilon
AGEB	236	9.2405
Private inhabited houses with radio.	8	9.2906
Private inhabited houses with washing machine.	8	9.2906
Neighborhood.	FOMERREY LA UNIDAD	9.4290
Private inhabited houses with fixed telephone line.	8	9.6649
Population 18 years old and over with post-basic education	8	9.8197
Female population from 18 years old and over with post-basic education.	8	9.8197
Neighborhood.	HACIENDAS DEL CANADA	9.9128
Neighborhood.	RIBERAS DE GIRASOLES	9.9128
Male population from 18 years old and over with post-basic education.	8	10.9389

Figure 4 shows a heat map for the incidence of business robbery, which is concentrated in the adjacent area of San Nicolas de los Garza municipality, the red areas describe shopping areas or industrial zones.

There exist several types of businesses that can share characteristics and others that can be unique for each business in particular, we did a drill-down to select a business subset.

In the country there is a large number of convenience stores, whether they are family or franchise stores (OXXO with 11,000 stores , SEVEN or EXTRA) [9], which work 24 hours a day, this exposes them to certain risk conditions of being stolen, but this condition is not the only variable that determines the possibility of theft. The total revenue of the year 2012 for the convenience stores in Mexico was 6,948.8 million, and has an estimated increase of 36% in this sector for the year 2017 [6] , which turns to this type of business highly relevant for the national economy. But like any other business, it is exposed to theft. According to the national criminal traffic light during the years 2012 and 2013, the business robbery in Nuevo Leon was over the national average [7] and in the criminal traffic lights of the municipality of General de Escobedo during the years 2012 and 2013 there was an increase over the historical average. [8] The gas stations share some features with the convenience stores, the most outstanding is that they are available 24 hours a day, in many cases a convenience store, OXXO

- SEVEN - EXTRA, is physically located in a gas station. What we aim is to classify the convenience stores and gas stations robberies by means of different characteristics; for example, the time of the crime, location, etc.

The variable with greater epsilon is the EARLY-MORNING Turn with 6.1941. The table 7 shows the four values for the variable turn, where the occurrence trend of crime type GSO is clearly seen at nights, increasing substantially in the early morning and down on the day and in the evening. This is most visible in table 8 where you can see how from 8 pm the epsilon has a positive sign and this trend is maintained until 6 am o'clock, We also see that the value of epsilon grows from 9 pm until 2 am, then it decreases, but maintains a positive sign up to 7 am. The time is an important variable for businesses as gas stations, OXXO - SEVEN - EXTRA since all these establishments work 24 hours.

Table 7. Epsilon turn.

Variable	Value	Epsilon
Turn	AFTERNOON	-4.6823
Turn	DAY	-1.9121
Turn	NIGHT	1.5137
Turn	EARLY MORNING	6.1941

Table 8. Epsilon Hour.

ONLY HOUR	0	1	2	3	4	5	6	7	8	9	10	11	12
Epsilon	2.64	3.60	3.37	2.17	2.22	1.14	2.16	0.56	-3.85	-0.60	-2.03	-0.92	-1.95

ONLY HOUR (cont.)	13	14	15	16	17	18	19	20	21	22	23
Epsilon	-3.83	-1.74	-2.51	-0.03	-2.29	0.09	-0.30	.55	.48	.87	1.81

4.3 Business Robbery Subset Model

In figure 5 there are some series of bars representing the number of incidences of type (GSO), for this case there were used very specific points due to we obtained the location (coordinates) from the establishments type GSO. As you can see the incidence is higher in those establishments which are located in important ways of communication as opposed to those found within the neighborhoods. For example, at the beltway intersection from Saltillo to Nuevo Laredo and the road to Colombia (see Figure 5 on the right) there are two establishments that have a total of 42 theft from 270 over a period of a year. This is because they offer a 24 hours service and its location has different escape routes coupled with they are settled on the periphery between the municipality and Monterrey city.

In the case of the gas station with a bigger number of theft, 12 robberies from 71, is highly insulated and the way in which the crime is committed is; when a vehicle is loading fuel and gets underway without paying it, this is encouraged by the isolated gas station because it is located in the beltway from Saltillo to Nuevo Laredo and there are only uninhabited lands.

4.4 House Robbery Model

For house robbery there are recorded a total of 419 crimes. In table 9 we observed a series of variables with higher epsilon, which outline house robbery. With these variables, we see a trend of home theft determined by the location and the socio-economic level where the dwelling is located.

Table 9. House robbery.

Variable	Value	Epsilon
Private inhabited houses with fixed telephone line.	7	6.3344
Private dwellings houses with internet.	7	6.3687
Private inhabited homes with electricity, piped water and drainage.	6	6.6232
Private inhabited houses with radio.	7	6.9621
Private inhabited houses with TV.	6	6.9644
Private inhabited houses with 3 bedrooms or more.	7	7.0361
Female population from 12 to 14 years old who do not attend school.	1	7.4953
Population from 18 years old and over with post-basic education.	7	7.5056
Private inhabited houses with a washing machine.	7	8,4
AGEB.	679	8.7419
Zip code.	66085	9.2692
Male population from 18 years old and over with post-basic education.	7	9.8030
Neighborhood.	PRADERAS DE SAN FRANCISCO	10.0007

With the help of Figure 6 see the geographical location of the areas with the highest incidence of r house robbery, mostly in the northeast, also it was located a small signal regarding the possible timetable of house robbery which is from 4 pm to 8 pm, but we have to keep in mind that the time is the moment in which it the crime was reported. Therefore it can be assumed that the reason for this time is that during this schedule is when people return to their houses and becomes aware of the offence, which also allows us to assume that the vast majority of house robbery is performed when there are no people in the housing.

5 Conclusions

In this paper we have modeled three different crime types using a Nave Bayes classifier using data taken from the municipality of General Escobedo in Nueva Leon. One of our chief conclusions is that in order to develop tools for decision support or intervention in crime in Mexico it is first necessary to develop a

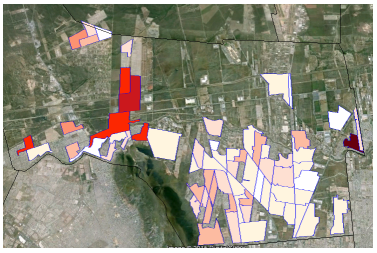

Fig. 3. Domestic violence

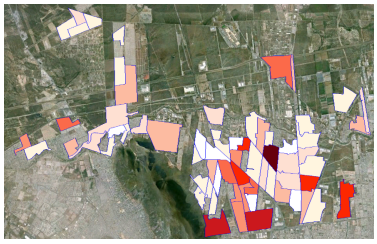

Fig. 4. Business robbery

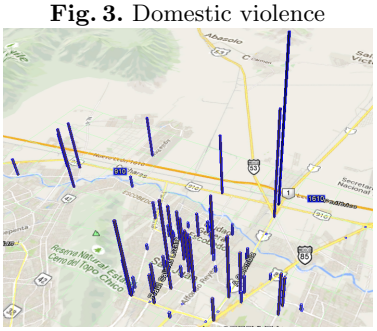

Fig. 5. Gas station and convenience store robberies

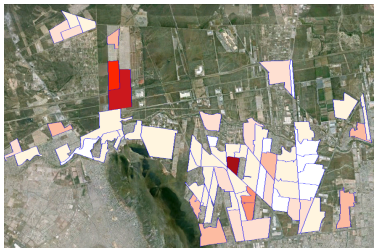

Fig. 6. House robbery

framework in which crime data can be reported and recorded in a faithful and timely fashion. Effectively, if we wish to predict crime we need to first plan to predict crime. There are several key elements associated with this notion: First of all, it is necessary to be able to record at the crime scene relevant characteristics of the crime. This could be simply done with a smart phone app. In particular, the crime should be georeferenced. Next the data should be transferred to a central data base from which any relevant analysis could be carried out.

Another important conclusion is that crime is complex. What we mean by that in the current context is twofold: First, that there exists an extremely large universe of potential predictors that are risk factors for crime and, secondly, that there is a rich hierarchy of crimes themselves each with its own predictive profile. For instance, as we saw in this paper, the profile for business robbery and the sub-group of convenience stores and gas stations share different characteristics but others are quite defining for the sub-group, the time variable is an example, even within the subset of the type GSO we can separate only stores or gas stations only and see how it affects the businesses location differently. How isolated an establishment is can result in an increased number of thefts, as this was the case for convenience stores that are on the periphery of occupied areas and are therefore quite isolated. The same applies to the gas stations with the largest number of robberies.

In spite of the complexity of the problem and the sparse data available we have shown that first of all a data mining approach to predict many different

types of crime is possible and that, indeed, there is a substantial degree of predictability; and secondly, that risk profiles for a crime type can be determined which can then lead to possible directed interventions by the authorities or local communities. In particular, by appending socio-economic and socio-demographic data from the Mexican census we saw that rich profiles associated with where the crimes took place could be generated for the different crimes types. Drill downs can be performed by neighborhood, zip code, AGEB, crime, it can even be done by each establishment type, but it will always be conditioned to the data quality.

We found that domestic violence in particular was associated with a very characteristic profile in terms of the socio-economic and socio-demographic characteristics of the areas where domestic violence events were more common. Among these were low educational achievement, presence of very small children, relatively large immigrant population and cramped living conditions. All these are social stress factors that together paint a coherent picture of the circumstances under which domestic violence is more likely. However, here we see their relative contributions as measured by the Nave Bayes score in a multifactorial setting. Of course, we must emphasise that here we are not necessarily identifying direct causal factors but, rather, statistically significant correlations.

References

1. Chen, H., Chung, W., Xu, J.J., Wang, G., Qin, Y., Chau, M.: Crime data mining: A general framework and some examples. Computer 37(4), 50–56 (Apr 2004), http://dx.doi.org/10.1109/MC.2004.1297301
2. Estivill-Castro, V., Lee, I.: Data mining techniques for autonomous exploration of large volumes of geo-referenced crime data. In: Proc. of the 6th international conference on geocomputation (2001)
3. Keyvanpour, M.R., Javideh, M., Ebrahimi, M.R.: Detecting and investigating crime by means of data mining: a general crime matching framework. Procedia Computer Science 3(0), 872 – 880 (2011), http://www.sciencedirect.com/science/article/pii/S1877050910005181, world Conference on Information Technology
4. McCue, C.: Data mining and predictive analysis: Intelligence gathering and crime analysis. Butterworth-Heinemann (2014)
5. Nath, S.V.: Crime pattern detection using data mining. In: Web Intelligence and Intelligent Agent Technology Workshops, 2006. WI-IAT 2006 Workshops. 2006 IEEE/WIC/ACM International Conference on. pp. 41–44 (Dec 2006)
6. Web: El economista (October 15th 2013), http://eleconomista.com.mx/industrias/2013/05/06/tiendas-conveniencia-copan-mercado-detallista
7. Web: RRS & Asociados S.C. (October 20th 2013), http://www.prominix.com/sblock/web/index.php?new=45.
8. Web: Semaforo delictivo (gobierno de Nuevo Leon) (May 18th 2013), http://www.semaforo.com.mx/
9. Web: Universal, Notimex: OXXO tienda ingreso (October 10th 2013), http://www.eluniversal.com.mx/finanzas-cartera/2013/oxxo-tienda-ingresos-936226.html.

10. Web: Instituto Nacional de Estadstica y Geografa (INEGI) (July 15th 2015), http://www.inegi.org.mx/sistemas/consulta resultados/ageb urb2010.aspx