Automatic Spatio-temporal Characterization of Criminal Activity. A New Algorithm and a Public Security Decision-support System

Víctor Manuel Martínez-Hernández[1], Salvador Godoy-Calderón[1], Francisco Hiram Calvo Castro[1] and Marco A. Moreno-Armendáriz[2]

[1] Artificial Intelligence Laboratory
[2] Pattern Recognition Laboratory
Research Computer Center CIC-IPN
victorm@cic.ipn.mx, sgodoyc@cic.ipn.mx,
hcalvo@cic.ipn.mx, marco_moreno@cic.ipn.mx

Abstract. A new algorithm, using inductive learning techniques and capable of finding criminal activity trends is presented. This algorithm inductively constructs spatio-temporal characterizations of each type of criminal activity previously registered in a given study region. The program that executes the algorithm is also presented, along with some discussion about the advantages of its use. This kind of characterization provides valuable support for the public-security decision making process. It allows decision makers to better plan surveillance patrols and to concentrate efforts and/or resources on those areas (and times) of highest criminal incidence.

Keywords: crime analysis, inductive learning, crime patterns, crime prevention, public security, etc.

1 Introduction

Nowadays public security and crime fighting activities are some of the most important social concerns in every large city around the world [5]. Despite the enormous quantities of human and financial resources that Governments devote to these activities, the need for alternate mechanisms to increase the effectiveness and efficiency of the police forces daily work remains evident [7]. One of the main variables that severely limits both the effectiveness and efficiency of those activities is the reaction time to crime occurrences. Even the greatest efforts in creating and managing special "immediate reaction" forces, have shown that the marginal gain in this issue always turns out to be insufficient. And indeed it is insufficient to reduce the general crime incidence and to substantially modify the insecurity sensation among the citizens.

A better perspective of the situation can be achieved if the problem is ported to the scope of prevention instead of that of reaction. If public security authorities could determine, with an acceptable degree of precision, when and where criminal activities are characteristically taking place, a double benefit would be achieved. On one hand, it would be possible to concentrate all the logistic activities and resources necessary to prevent that specific kind of criminal activity in the geographic area and time frame

© G. Sidorov, M. Aldape, M. Martínez, S. Torres. (Eds.)
Advances in Computer Science and Engineering.
Research in Computing Science 42, 2009, pp. 103-111

Received 19/03/09
Accepted 27/04/09
Final version 07/05/09

forecasted. On the other hand, it would become possible to establish, in a dynamic but well supported manner, many of the parameters for public security daily activities, such as the design of specific patrolling itineraries, the allocation of resources and, of course the scheduling of security operations and of information and prevention campaigns by mass media.

As an additional asset for public security officials and politicians, the capability to learn the spatio-temporal trend in crime activities would provide an adequate foundation for evaluating their performance in decision making [7]. This foundation would be provided by the simple comparison between the foreseen trends and the countermeasures adopted and/or the outcome of the undertaken operations. All this would allow for a better determination of the annual budget devoted to public security, as well as its better allocation to the various entries related to this complex activity.

In this paper we present a new model and a strategy of criminal activity predicting by discovering its spatio-temporal trends within a specific time and zone under study [2]. The specific strategy proposed is based on an inductive learning process commonly used in pattern recognition for classification and clustering problems [1] [3]. The aforementioned model is one of the first steps in the process of building an intelligent automatic system to support decision-making processes in public security affairs.

2 Proposed Methodology

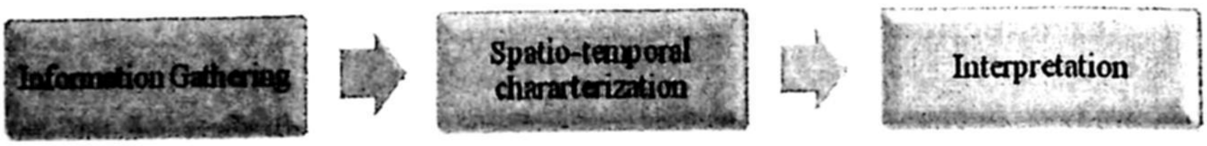

Fig. 1. Proposed Methodology

2.1 Information Gathering

It is necessary to gather information on crimes that occur in the region under study. The common sources of information are:

- Reports of Public Prosecutions.
- Citizen Alerts.

These reports are analyzed with the intention that each registered crime is transformed into a crime pattern, consisting of three components:

$$D = \{Crime\ Act,\ Space,\ Time\}$$

- Crime Act. Specifies the type of criminal activity that is registered (theft, rape, murder, etc.).

- Space. Specifies the place where the criminal act took place (surveillance sector, colony, street, etc.)
- Time. Indicates the time at which the crime occurred (year, month, day, hour, etc.).

2.2 Spatio-temporal Characterization

The learning process involves three specific phases: Aggregate crime patters into crime families, constructing two inductive definitions, one positive and one negative, for each crime family and, finally, constructing a neutral inductive definition for all types of criminal activity. Following, each of these stages is described:

2.2.1 Aggregate Crime Patterns into Crime Families

Once all registered crimes are converted into a crime patterns, the next step is to construct a classification of reference. To construct such sample, all patterns are clustered by the kind of preventive measures, material and logistic resources needed to prevent them. The choice of the number and nature of these families (classes) is entirely dependent on local policy and strategies. At the end of this stage, each class' patterns are ordered in time and space. Table 1 shows the resulting structure:

Table 1. classification of Reference

$D_1 = (\{Space\}_1, \{Time\}_1)$	
. . .	Crime Family #1
$D_p = (\{Space\}_p, \{Time\}_p)$	
. . .	. . .
$D_q = (\{Space\}_q, \{Time\}_q)$	
. . .	Crime Family #k
$D_n = (\{Space\}_n, \{Time\}_n)$	

2.2.2 Constructing Inductive Definitions

At this stage it is necessary to determine the subsets of pattern features that will constitute the spatio-temporal properties to look for. A property is a subset of pattern features associated with specific values including both a spatial and temporal component:

$$P_i = \left\{ \begin{matrix} Property_{i_1} \\ < Value_{i_1} \end{matrix}, \ldots, \begin{matrix} Property_{i_n} \\ < Value_n \end{matrix} \right\}$$

Once these properties are defined, a query, looking for those properties, is run over the classification of reference. Those properties that appear more commonly in a class and less in others are considered to be characteristic of the class. The union of all the characteristic properties of a class (C_i) is called an Inductive Definition of that class, and has the following structure:

$$C_i = \{Property_{i_1}\} \cup \{Property_2\} \cup \ldots \cup \{Property_{i_m}\}$$

The algorithm includes the construction of two inductive definitions for each class and a neutral for all them. These definitions are constructed as follows:

a) **Positive Inductive Definition (C_i^+).**

1. Determine the characterization factor (F) for each spatio-temporal property (P) in each class (C_i):

$$F_{C_i}^{+}(P) = \frac{nP_{C_i}}{nP} \qquad 0 \leq F_{C_i}^{+}(P) \leq 1 \qquad (1)$$

Where nP_{C_i} is the number of occurrences of property P in class C_i, and nP is the total number of occurrences of property P in all classification of reference.

2. All spatio-temporal properties with characterization factor greater than a threshold β_1 are considered to be included in the positive inductive definition for the class. The more close $F_{C_i}^{+}(P)$ is to 1, the more characteristic P is in class C_i.

b) **Negative Inductive Definition (C_i^-).**

All spatio-temporal properties with characterization factor less than a threshold β_2 and appearing in β_2' percent of the classes are included in the negative inductive definition for the class.

Threshold β_2 indicates the maximum characterization factor value a property must have to be considered as a negative characteristic. Threshold β_2' defines a minimum percentage of classes in which the property must appear to be considered as negative characteristic.

2.2.3 Building a Neutral Inductive definition.

a. Calculate the average characterization factor ($\hat{S}$) of each property (P) in all classes:

$$\hat{S}(P) = \frac{1}{k}\sum_{i=1}^{k} F_{C_i}^{+}(P) \qquad (2)$$

Where k is the number of classes.

b. For a property to be included in the neutral inductive definition it must satisfy the following condition:

P should have $|F_{C_i}^{+}(P) - \hat{F}(P)| \leq \beta_i$ in at least β_i' % of classes.

Once the spatio-temporal characterization is finished, positive, negative and neutral inductive definitions are integrated in the following manner:

Table 2. Inductive definitions

$C_i^{+} = \{Property_{i_1}\} \cup \ldots \cup \{Property_{i_{m_{i_1}}}\}$ $C_i^{-} = \{Property_{i_1}\} \cup \ldots \cup \{Property_{i_{m_{i_k}}}\}$	*Criminal Family #1*
⋮	⋮
$C_k^{+} = \{Property_{k_1}\} \cup \ldots \cup \{Property_{k_{m_{k_1}}}\}$ $C_k^{-} = \{Property_{k_1}\} \cup \ldots \cup \{Property_{k_{m_{k_k}}}\}$	*Criminal Family #k*
$L^{0} = \{Property_1\} \cup \ldots \cup \{Property_r\}$	

2.3 Interpretation

The set of inductive definitions have the following semantic:

a) The positive inductive definition of a crime family identifies places and moments where that specific criminal activity characteristically occurs (hotspot), therefore it suggests when and where it is necessary to increase surveillance and patrolling activities.

b) The negative inductive definition of a crime family identifies places and moments where that specific criminal activity characteristically does not occur, so, suggesting to re-assign resources from those places and moments to those hotspots identified in (a).

c) The neutral inductive definition has great importance in crime analysis; its properties identify those places and moments where the great majority of criminal activities take place although not in a characteristic manner. This definition suggests where and when it is necessary to implement more intensive surveillance and maybe to start preventive and informative mass media campaigns.

3 Experiments and Results

To test the proposed algorithm a set containing 16,995 registered crimes was used. This dataset was provided by the District of Cuautitlan Izcalli, Mexico and all crimes were committed in 195 neighborhoods in that district over a period of time from January 2004 to December 2008.

The registered crimes were converted into crime patterns and cluster into 17 crime families as shown in Table 3:

Table 4 shows the threshold used in the experimentation.

Table 3. classes and patrons used in the experimentation

Crime	Patterns
Housebreaking	393
Damage to private property	2635
Drug dealing	90
Illegal weapons	172
Homicide	435
Injuries	4890
Domestic violence	720
Burglary	477
Commerce and industry burglary	901
Robbery	2675
Assault	718
Public Transport mugging	84
Car parts theft	276
Smuggling	98
Car theft	2295
Kidnapping	89
Raping	47

Table 4. Threshold used in the experimentation.

Threshold	Value
β_1	0.65
β_2	0.2
β_2'	0.75
β_3	0.05
β_3'	0.30

Table 5 shows the number of positive and negative characteristic properties found for each crime family and the number of neighborhoods in which those criminal activities are characteristic.

Table 5. Positive and negative characteristic properties found for each crime family

Crime Family	Positive discriminating characteristic properties	Negative discriminating characteristic properties	Number of neighborhoods in which criminal activities are characteristic
Housebreaking	31	15	20/195
Properties Damage	263	5	73/195
Drugs dealing	18	25	4/195
Illegal Weapons	14	32	6/195
Homicide	49	8	28/195
Injuries	530	16	95/195
Domestic violence	56	12	38/195
Burglary	33	11	23/195
Industry burglary	56	10	36/195
Robbery	211	5	68/195
Assault	49	8	33/195
Transportation mugging	23	13	3/195
Car parts theft	46	11	20/195
Smuggling	24	2	7/195
Car theft	220	22	77/195
Kidnapping	8	2	6/195
Raping	4	4	4/195

The table 6 shows some spatio-temporal properties affected by Robbery.

Table 6. Characteristic properties affected by Robbery

Neighborhood	Month	Time
Bellavista	May	(15:01-18:00 hr)
	November	(15:01-18:00 hr)
Bosques de la Hacienda	June	(00:01-03:00 hr)
	September	(12:01-15:00 hr)
Bosques de Morelos	May	(09:01-12:00 hr)
	November	(09:01-12:00 hr)
Cofradia II	February	(21:01-00:00 hr)
Colonias del lago I	September	(00:01-03:00 hr)

Table 7 shows some spatio-temporal properties affected by some other crime families.

Table 7. Spatio-temporal properties affected by crime families

Neighborhood	Month	Time	Crime Families
Cumbria	January	(21:01-00:00 hr)	House Burglary Homicide Street Robbery Transportation mugging Car parts theft Car theft

4 Conclusions

The development of a new crime analysis model, based on inductive learning techniques is presented. The proposed model has some advantages over traditional criminal analysis: To mention some of them, algorithms of clustering are used in this kind of analysis in order to find those areas which are affected by a kind of criminal activity in an period of time (year, month, week, etc), so it is possible to use layers in geographical information systems (GIS) to represent the result of this analysis. However, criminal activities have to be analyzed by separated, so there is not a direct way to determine which criminal activity have more impact in areas with several criminal activities; the relation between criminal activities in an analyzed area is very important in order to implement preventive plans that reduce those criminal activities with more impact. The proposed model used characteristic spatio-temporal properties to decide which criminal activity have more impact in the analyzed area.

In the same context, traditional criminal analysis do not mention anything about how to use their result, when and where it is necessary to increase surveillance and patrolling activities, where and where it is possible to reduce surveillance and resource to attend those area which requires more resources, and finally which areas require special attention, because the great majority of criminal activities take place although not in a characteristic manner.

Other advantage of the proposed model is that the inductive definitions of each class, constructed by the spatio-temporal characterization procedure constitutes, by itself, valuable information for describing the criminal trends being studied; on the other hand, by adjusting the value of each threshold, it is possible to fine tune the level of precision we want to have in the inductive description of each class.

The type of inductive learning procedure presented in this paper is commonly used for supervised classification [1, 3] and conceptual clustering [5]. The proposed analysis method uses it in order to discover crime trends in large databases.

Finally, it is noteworthy that the proposed model opens new possibilities for the design and implementation of crime analysis systems where traditionally only statistical analysis had been used.

Reference

1. Baskakova L.V., Zhuravliov Y.I., "Modelo de algoritmos de Reconocimiento con Conjuntos de Representantes y sistemas de Conjuntos de Apoyo". Zh. Vivhislitielnoi Matematiki I Matematichskoi Fiziki. Tom 21, No.5, URSS, 1981.
2. Brachman R.J., Anand T., "The process of knowledge discovery in databases: A human centered approach". In U.M. Fayyad, G. Piatetsky-Shapiro, P.Smyth and R. Uthurusamy (editors). Advances in Knowledge Discovery and Data Mining, Chap.2. AAAI/MIT Press, 1996.
3. Diukova E.V., "About a parametric model of KORA-type Recognition algorithms". Soovshenia po prikladmoi matematiki. URSS, 1983.
4. Michalski R., Steep R.E. "A theory and Methodology of Inductive Learning". J. Carbonell, editor. Machine Learning: A Artificial Intelligence Approach. Chapter 11, Ed. Tioga, Palo Alto, California. 1984.
5. Oakley B. Robert, Goldberg M. Eliot, Dziedzic J. Michael, "Policing the New World Disorder: Peace Operations and Public Security". University Press of the Pacific. June 2002.
6. Quinlan J.R., "Discovering Rules by Induction from Large Collection of Examples". Michele, D. Editor, Expert System in the Microelectronics Age. Edinburgh University Press. 1979.
7. United Nations Publications. "Fighting crime in the twenty-first century: five regional meetings prepare for 1990 congress". 1990. Eighth UN Congress on the Prevention of Crime and the Treatment of Offenders. UN Chronicle. July 28, 2005.