The CR-Ω+ Classification Algorithm for Spatio-Temporal Prediction of Criminal Activity ♦

S. Godoy-Calderón[2], H. Calvo*[1,2], V. M. Martínez-Hernández[2], M. A. Moreno-Armendáriz[2]

[1]Nara Institute of Science and Technology, Takayama, Ikoma, Nara 630-0192, Japan
sgodoyc@cic.ipn.mx, *hcalvo@cic.ipn.mx, *calvo@is.naist.jp
mhernandezb07@sagitario.cic.ipn.mx, marco_moreno@cic.ipn.mx
[2]Centro de Investigación en Computación CIC-IPN
Av. Juan de Dios Bátiz s/n esq. Manuel Othón de Mendizábal, 07738, Mexico City, Mexico

ABSTRACT
We present a spatio-temporal prediction model that allows forecasting of the criminal activity behavior in a particular region by using supervised classification. The degree of membership of each pattern is interpreted as the forecasted increase or decrease in the criminal activity for the specified time and location. The proposed forecasting model (CR-Ω+) is based on the family of Kora-Ω Logical-Combinatorial algorithms operating on large data volumes from several heterogeneous sources using an inductive learning process. We propose several modifications to the original algorithms by Bongard and Baskakova and Zhuravlëv which improve the prediction performance on the studied dataset of criminal activity. We perform two analyses: punctual prediction and tendency analysis, which show that it is possible to predict punctually one of four crimes to be perpetrated (crime family, in a specific space and time), and 66% of effectiveness in the prediction of the place of crime, despite of the noise of the dataset. The tendency analysis yielded an STRMSE (Spatio-Temporal RMSE) of less than 1.0.

Keywords: Logical-combinatorial pattern recognition, forecasting models, supervised classification, crime analysis.

RESUMEN
Presentamos un modelo de predicción espacio-temporal que permite la predicción de la actividad criminal dentro de la región estudiada usando clasificación supervisada. El grado de pertenencia de cada patrón es interpretado como el incremento o decremento previsto de la actividad criminal para un tiempo y lugar específico. El modelo propuesto de predicción CR-Ω+ está basado en la familia de los algoritmos Lógico-Combinatorios Kora-Ω para clasificación supervisada. Estos operan sobre volúmenes grandes de datos obtenidos a partir de fuentes heterogéneas de información, con un proceso inductivo de aprendizaje. Proponemos diversas modificaciones al algoritmo original de Bongard, así como el de Baskakova y Zhuravlëv, las cuales mejoran el desempeño de la predicción en el conjunto de datos estudiados de actividad criminal. Realizamos dos análisis: predicción puntual y análisis de tendencias, los cuales muestran que es posible predecir puntualmente uno de cuatro crímenes a ser perpetrados (por familia del crimen) en un tiempo y espacio específicos, así como un 66% de predicción del lugar del crimen, a pesar del ruido en el conjunto de datos de entrada. El análisis de tendencias dio como resultado un RMSE Espacio-Temporal (STRMSE) menor a 1.0.

1. Introduction

Public security and crime fighting activities are some of the most important concerns of both citizens and government. Large amounts of money, human resources, quipment and services are devoted to these activities. In addition, there is a constant concern to justify budget allocated for police forces. Nevertheless, experience has shown that despite the resources devoted to these activities, it never seems to be enough [14, 18, and 22]. If the police forces could anticipate, with an acceptable degree of precision, when and where criminal activities of a specific kind are going to

♦ We thank the support of the Mexican Government (SNI, SIP-IPN, COFAA-IPN, and PIFI-IPN), CONACYT, the Japanese Government and the President of the Municipality of Cuautitlan Izcalli, David Ulises Guzmán Palma, for their support and help in obtaining the data used in this research. The second author is currently a JSPS fellow.

take place, it would achieve a double benefit [8, 21]. First, it would be possible to concentrate the necessary logistic activities and resources fighting that specific kind of criminal activity in the geographic area and forecasted time frame, and the comparison between the amount of resources allocated to police forces and the results achieved by them may result in a more adequate basis for planning and distribution of public security budgets.

There are several works devoted to the study of spatial and temporal decisions made by criminals, i.e., identifying hotspots where criminal activity is concentrated—see [1, 2, 6, 17, 23, 24, 25, 26, 27, and 30]. A widely used method is the Spatial and Temporal Analysis of the Crime Program (STAC) [4], which clusters crime points within ellipses [3]. Jefferis [16] surveys additional hotspot methods, the most sophisticated of which employs the kernel density estimation method [19]. Nevertheless, the main disadvantage of statistical methods is that they do not offer additional semantic information for describing the phenomenon under study. In the specific case of crime prediction, this kind of information is highly desirable, as it is needed to support decision-making processes and, in general, to prepare preventive and corrective policies. Because of this, we have selected inductive classification methods over statistical ones in order to generate an inductive description of each type of criminal activity studied. These descriptions by themselves constitute valuable information that provides a general overview of the criminal activity scenario. Further, by using these inductive definitions, it is possible to identify the expected increase or decrease in specific criminal activities that will most likely occur in specific geographic areas and times.

This paper reports the design, experimentation and results achieved by the proposed model of criminal activity forecasting within a specific time period and location using several different supervised classification techniques. In Section 2, we present details of our forecasting model, from general forecasting with Inductive Supervised Classification (§2.1) to our particular implementations of CR-Ω+ (§2.2) and our variant with general discrimination (CR-Ω+M) (§2.3). Then in Section 3, we show the results of our experiments. We perform two analyses: punctual hotspot prediction (§3.1), and tendency analysis (§0). For the first analysis, we use data from the municipality of Cuautitlan Izcalli, State of Mexico. Within this analysis we perform experiments for spatial and temporal location of crime (§3.2) and expected family of crime (§3.2.2). For the tendency analysis (§3.3), we use data from the Sacramento, California (CA), Police Department. Finally, in Section 4, we draw our conclusions and discuss future work.

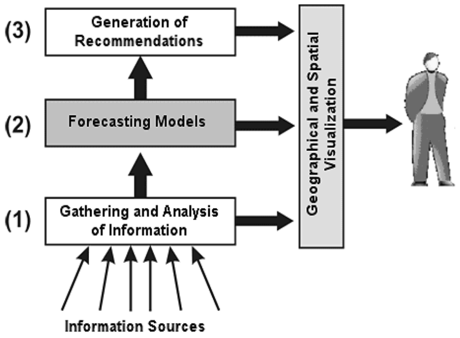
(3)
Generation of Recommendations
(2)
Forecasting Models
(1)
Gathering and Analysis of Information
Information Sources
Geographical and Spatial Visualization

Figure 1. General architecture of the computerized system to support decision-making processes in public security.

2. Forecasting Model

The forecasting model reported here is part of a larger system designed to prevent and react to crime. This larger system is made-up by three main layers as shown in Figure 1.

The function of the first layer is to gather, standardize and analyze data from each of six established information sources. In terms of pattern recognition, layer one (in Figure 1) constitutes the supervision sample. The second layer contains several prediction algorithms—this paper focuses on this layer. The input for the third layer is the set of predictions made by the algorithms of the previous layer and it generates recommendations for addressing the forecasted scenarios. The generated recommendations are classified into four types: (1) recommendations on patrol logistics covered by police vans or walking officers, (2) recommendations to begin specific preventive campaigns through the mass media, (3) recommendations on the implementation of corrective operations, and (4) general recommendations on the state of the police force. The forecasting model described herein is part of the second layer shown in Figure 1.

Careful structuring of the input data is of utmost importance, in order that the forecasting model can be efficient and has an acceptable level of precision [9]. The problem of crime prediction requires several different information sources, all of them directly related with public security, but not easily accessible. For this project, we have chosen six information sources arranged into four categories as follows: information on (1) crimes committed, (2) citizens' reports, (3) resources and police activities, and (4) socioeconomic data of the region under study. In each of these categories, information should be precisely located in time and space.

2. 1 Forecasting with Inductive Supervised Classification

One of the most interesting tasks of the pattern recognition discipline is the study of forecasting models [8, 21]. The forecasting problem can be treated as a classification problem; this allows taking advantage of the large number of available classification algorithms. The great majority of current forecasting models has a statistical nature and is mainly devoted to time series analysis (for an exhaustive review of these methods, refer to [8] and [13]). As stated previously, we are interested in the particular semantics of crime analysis, so we propose an inductive forecasting model.

The forecasting model is expressed as a supervised classification problem in the following terms. Given a database containing a set of patterns corresponding to crimes perpetrated within the region under study that are spatio-temporally labeled, group such patterns into crime families, where a crime family consists of all the crime patterns corresponding to similar crimes that are fought with the same resources. These families form the supervision sample to be used by the classification algorithms. Afterwards, a learning process is carried out in order to describe each family in both positive and negative ways. In order to predict a specific criminal scenario (i.e., the time, location and type of criminal activity to be predicted), a pattern containing all the relevant data is constructed and submitted for classification in accordance with the supervision sample previously assembled. The classification algorithm gives as a result the degree of membership of such a pattern to each one of the established families. Consequently, each degree of membership is interpreted as the forecasted increase or decrease in the criminal activity for the specified time and location.

In order to make forecasts, a careful design of the classification problem semantics is required. Specifically, this includes three basic aspects of the problem: (1) the objects under study and the attributes or features that will be used to describe them; (2) the number of classes and how patterns will be classified; and (3) what kind of learning the classification algorithm will use. Each one of these aspects is discussed below.

2.1.1 Objects and Descriptive Features

The objects under study are either criminal activity scenarios or citizens' reports recorded in a given geographical area at a given time. Temporal information includes the date and time components, whilst the location refers to the area in which the criminal activity was recorded. For each one of the studied scenarios an r-dimension pattern is defined containing all the data available. The resulting set of these patterns forms the supervision sample.

We perform two different modes for analyzing the criminal data, considering two different semantics for the crime families to be identified.

Learning Stage:

1) For each class in the supervision sample, build the following inductive definitions:

- ➢ $\theta_1^+(C_i)$ positive description of the most representative patterns in the class. (feature set present at least β_1^+ times in C_i and no more than $\beta_1'^+$ times in any other class C_j with $j \neq i$)
- ➢ $\theta_1^-(C_i)$ negative description of the class. (feature set present at least β_1^- times in any other class C_j with $j \neq i$ and no more than $\beta_1'^-$ times in C_i)

2) Calculate the inductive definition of the neutral properties common to all the classes θ^0 (feature set appears at least β^0 times in all classes).

Re-Learning Stage:

3) For each class in the supervision sample, build the following inductive definition:

- ➢ The $\theta_2^+(C_i)$ positive description of other less representative patterns in the class. (feature set present at least β_2^+ times in C_i and no more than $\beta_2'^+$ times in any other class C_j with $j \neq i$)

Classification Stage:

4) For each one of the patterns to be classified (requested forecast), record the positive characteristic, the negative characteristic and the complementary properties that it fulfils.

5) Apply a solution rule that indicates the membership of the classified pattern to each of the defined classes. For this case,

$$S(C_i,P)=\left|F_P\cap\theta_1^+(C_i)\right|-\left|F_P\cap\theta_1^-(C_i)\right|+\frac{1}{2}\left|F_P\cap\theta_2^+(C_i)\right|$$

where F_P represents the set of all feature subsets from the pattern to be classified.

Figure 2. The CR-Ω+ Algorithm

2.1.2 Classes

The first mode groups the supervision sample patterns into three classes that indicate the kind of environment in which the criminal activity was recorded. These three classes represent criminal activities committed in

Mode 1. (1) Public roads and highways, (2) homes , (3) stores and shops.

The second mode groups the patterns into four classes based on the kind of social impact that the criminal activity has. These four classes are

Mode 2. (1) Robbery in all of its modalities, (2) homicide, (3) injury, (4) property damage.

These groupings were selected considering the main goal of the project, which is to provide recommendations to the public security authorities. Based on the recommendations of our system, authorities will have information to spawn direct actions to prevent the crimes perpetrated in public streets, shops, homes, or industries. The corresponding action could be, for example, to generate preventative campaigns

2.1.3 Learning

We use information regarding the crimes committed and reported to the Public Prosecutor's Office. The data corresponding to the crimes committed during a certain time can be used to test the proposed forecasting model. After having initially debugged the information provided by the Public Prosecutor's Office, a supervision sample containing the space-time location of the crimes committed is built. A classifier learning process is trained on data containing a good distribution of crime patterns across the three or four classes mentioned above, depending on the mode under investigation. This learning process consists of constructing inductive definitions that describe in positive and negative ways the patterns contained in each one of the classes in the supervision sample. The inductive definition, θ, corresponding to each C_i class, is an expression of the form:

$$\theta(C_i) = \bigcup P_m(C_i) \tag{1}$$

where each is a property identified among the patterns pertaining to the class [5]. The properties are subsets of descriptive features associated with specific values. For each class, a positive and a negative description are made. Once the descriptions are obtained, we can use an inductive classification algorithm.

2.2 The CR--Ω+Classification Algorithm

We propose a combination of the *Kora*-Ω algorithm which is an extension of the KORA-3 [5, 12,11, 10] algorithm and the *Representative Sets* (CR+) algorithm [3, 7], which both share the notion of property in the form of a subset of features associated with specific values in these features. A P_m property identified in the C_i class has the form shown in Eq.(2).

$$P_m = \left\lfloor \begin{matrix} x_p, & \ldots, & x_q \\ \langle v_p \rangle, & \ldots, & \langle v_q \rangle \end{matrix} \right\rfloor \tag{2}$$

Where $x_i \quad i = p..q$ are features used to describe the objects under study and each $\langle v_j \rangle \quad j = p..q$ is a specific value in the domain of the x_j feature observed among the patterns in the C_i class.

A property of this kind is called *a Characteristic Property* in the class if and only if it appears β_1^+ times among the patterns of the class and $\beta_1'^+$ times or less among the patterns of the complement of the class. A property is called *Negative in the Class* if it occurs at least β_1^- times in any other class, $\beta_1'^-$ and in the class. $\beta_1'^-$ is always smaller than β_1^+ If a property is found to repeat itself a number of times amongst the patterns of all the classes in the sample, it is called a *Neutral Property* and we use a β_0 threshold for that purpose. See Figure 2 for additional information.

Once the characteristic properties have been evaluated, the *complementary properties* are evaluated by executing the algorithm again with the characteristic and negative properties removed and using the threshold for this second stage: β_2^+ and $\beta_2'^+$.

The concept and associated procedures are taken from the *Kora*-Ω algorithm in order to describe both the most representative patterns of each class and the less representative ones. From the *CR+* algorithm, concepts for the negative description of each class and for the description of those features common to all classes (neutral properties) are taken. To find out specific details regarding both algorithms, readers are referred to [21] and [29]. The specific combination of these algorithms results in a new algorithm called the previously described *CR-Ω+ algorithm*.

The solution rule applied in the fifth step (see Fig. 1) includes determination of the relevance of the properties contained in an inductive definition to each defined class. This may be calculated in different ways. Typically, the weighing process is associated with an information weight for each feature in the property and the number of repetitions of the property amongst the elements of a class [12].

2.2.1 Example

Consider the following table:

TIME ID	MONTH	ZONE	C_1	C_2	C_3
8	July	**Santiago Tepalcapa**	1	0	0
8	May	**Santiago Tepalcapa**	1	0	0
8	May	**Santiago Tepalcapa**	1	0	0
1	January	Sección Parques	1	0	0
2	January	Sección Parques	1	0	0
4	April	Bosques del Lago	0	1	0
6	April	Bosques del Lago	0	1	0
8	May	Centro Urbano	0	1	0
5	July	Ensueños	0	1	0
5	July	Santiago Tepalcapa	0	1	0
6	July	**Centro Urbano**	0	0	1
6	June	**Centro Urbano**	0	0	1
6	March	**Centro Urbano**	0	0	1
5	February	Cumbria	0	0	1
4	June	Cumbria	0	0	1

We want to classify the following pattern:

P_1 =[8, February, Santiago Tepalcapa]

Let us use for this example the following thresholds: $\beta_1^+ = 3, \beta_1'^+ = 1 \beta_1^- = 3$ and $\beta_1'^- = 1$. This means that a feature set will be characteristic positive if it appears at least three times for this class, and it appears at most once for other classes. A feature set will be complementary if it appears at most once for this class and three times or more in other classes. Note that for this example we have chosen $\beta_1^+ = \beta_1^-$ and $\beta_1'^- = \beta_1'^+$. Although this is a common practice, it is not required by the algorithm.

A feature set can be composed of any combination of features. For this example, the following combinations are evaluated: [QUADRANT, DATE], [QUADRANT, ZONE], [DATE, ZONE], [QUADRANT, DATE, ZONE]. For this example we discarded the combinations which involve only one feature. Thus, it can be seen that the feature set [8, Santiago Tepalcapa] appears three times for C_1, and zero times in any other class so it is a positive characteristic for C_1. [6, Centro Urbano] appears three times for C_3, and zero times for the other classes, which means that it is a negative characteristic for C_1, and a positive characteristic for C_3. In addition, [8, Santiago Tepalcapa] appears three times in another class (C_1), so it is a negative characteristic for C_3. There are no other features repeated at least three times for this example. In this example there are no feature sets composed of three features [QUADRANT, DATE, ZONE] repeated at least three times.

We have calculated the positive characteristic features $\theta_1^+(C_i)$, and the negative features $\theta_1^-(C_i)$ for each class C_i :

$\theta_1^+(C_1)$={[8, Santiago Tepalcapa]} $\theta_1^-(C_1)$={[6, Centro Urbano]} $\theta_1^+(C_2)$ ={} $\theta_1^-(C_2)$ ={[8, Santiago Tepalcapa], [6, Centro Urbano]}
$\theta_1^+(C_3)$={[6, Centro Urbano]} $\theta_1^-(C_3)$={[8, Santiago Tepalcapa]}

Next, we proceed to calculate the complementary features $\theta_2^+(C_i)$ The complementary features are calculated like the positive features $\theta_1^+(C_i)$, but with different thresholds, namely $\beta_2^+ = 2$ and $\beta_2'^+ = 1$ This means that all feature sets that appear at least two times in this class, and at most once in another class will be complementary features.

$\theta_2^+(C_1) =$ {[8, May, Santiago Tepalcapa], [8,May], [May, Santiago Tepalcapa],
[8, Santiago Tepalcapa], [January, Sección Parques]}
$\theta_2^+(C_2) =$ {[April, Bosques del Lago], [5, July]}
$\theta_2^+(C_3) =$ {[6, Centro Urbano]}

Let us say we want to classify the pattern P_1=[8, February, Santiago Tepalcapa]. We calculate F_{P_1}, the set of all the subset features from P_1 and intersect with the characteristic and complementary sets. For example, its intersection with $\theta_2^+(C_1)$ yields {[8, Santiago Tepalcapa]}. Thus, $\left|F_{P_1} \cap \theta_2^+(C_1)\right|$=1.

The correspondence of intersections with characteristic, negative or complementary features is shown below:

$F_{P_1} \cap$	$\theta_1^+(C_1)$	$\theta_1^-(C_1)$	$\theta_2^+(C_1)$	$\theta_1^+(C_2)$	$\theta_1^-(C_2)$	$\theta_2^+(C_2)$	$\theta_1^+(C_3)$	$\theta_1^-(C_3)$	$\theta_2^+(C_3)$
$\lvert\ \rvert$=	1	0	1	0	1	0	0	1	0

Then we apply the solution rule shown in Figure 2 to determine the belonging of P_1 to each class.

C_i	$\theta_1^+(C_i)$	$\theta_1^-(C_i)$	$\theta_2^+(C_i)$	$S(C_i, P_1)$
C_1	1	0	(0.5) 1	1.5
C_2	0	-1	(0.5) 0	-1
C_3	0	-1	(0.5) 1	-.5

We conclude that pattern P_1 belongs to C_1 .

2.3.CR-Ω+ with General Discrimination (CR-Ω+M)

The CR-Ω+*Modified* Algorithm (henceforth referred as CR-Ω+M) is based on the previous algorithm, modifying the counting of features present in other classes, i.e., those related with the $\beta_1'^+$ and $\beta_1'^-$ thresholds.

For the previous algorithm, $\theta_1^+(C_i)$ was calculated by counting a feature set present at least β_1^+ times in C_i and no more than $\beta_1'^+$ times in any other class C_j with $j \neq i$. For the *CR*-Ω+M algorithm, we modify this last part, now requiring "no more than $\beta_1'^+$ times in the union of other classes C_j with $j \neq i$ ", that is, $\bigcup C_j, j \neq i$. The same occurs for calculating $\theta_1^-(C_i)$ requiring now that the feature set be present at least β_1^- times in $\bigcup C_j, j \neq i$ and no more than times in .

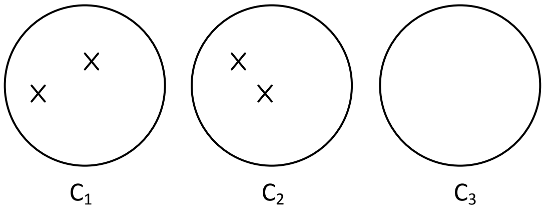

Figure 3. The *CR*-Ω+M classification algorithm will consider the X features as negative for class three with β_1^-=3, whereas the *CR*-Ω+ algorithm will not.

In Figure 3, we exemplify the effect of this modification. Given a threshold =3 for C3, feature X in the first algorithm (*CR*-Ω+) $\theta_1^-(C_3)$ would be empty, whereas for the *CR*-Ω+Malgorithm, the cardinality of the features would be 4, yielding $\theta_1^-(C_3)$={X}.

For the example from the previous section, $\theta_1^-(C_3)$={[8, Santiago Tepalcapa]}, the *CR*-Ω+ modified algorithm would add [8, May]: $\theta_1^-(C_3)$ ={[8, Santiago Tepalcapa], [8, May]}.

In the following section, we perform several experiments with real data to test the performance of the proposed algorithm.

3.Experiments

The input data should be complete, and there should be mechanisms to capture it systematically in due time and proper form [29]. In the Cuautitlán Izcalli district, located in Mexico, the local government launched the *Centro de Emergencias Cuautitlán* (CERCA, Cuautitlán Emergency Center) in 2007. An important part of its function is the gathering of the information corresponding to the three first categories of information sources. Therefore, this district was selected as a case study and test field for both the forecasting model reported herein and the computerized system for supporting the decision-making process on public security. We performed two analyses: punctual hotspot prediction (§3.1), and tendency analysis (§3.3). For the first analysis we used data from the municipality of Cuautitlán Izcalli, State of Mexico. Within this analysis we performed experiments for spatial and temporal location of crime (§3.2) and expected a family of crime (§3.2.2). For the tendency analysis (§3.3) we additionally used data from the Sacramento, CA, Police Department for validation and comparison.

3.1 First Analysis: Punctual Hotspot prediction

In this section, we describe the details of our experimentation for applying the different algorithms presented in previous sections to the two modes mentioned in Section 2.1.2. All experiments in this section use data from January 1st, 2007 to July 31st, 2007 from the municipality of Cuautitlán Izcalli, State of Mexico. In total, there were 1551 records such as those shown in Table 1. Not every record from the original sample was useful due to incompleteness or ambiguity when classified on the chosen relevant classes; for many records there was no precise information about the place where the crime was perpetrated.

3.1.1 Data Preprocessing

An example of the contents of the reports from the Public Prosecutor’s Office is shown in Table 1. For each reported crime, its properties are time, date, zone and which kind of crime was committed.

To smooth the first column in Table 1, time information was grouped in eight regions called quadrants. With this division, we split the twenty-four hours in a day in eight periods of three hours each. The corresponding times are shown in Table 1.

Time	Date	Residential Zone	Crime Class
22:30	Jan 15	Arcos del Alba	Stolen Vehicle
11:40	Apr 18	Atlanta	Stolen Vehicle
00:30	Jun 11	Bosques de la Hacienda	Familiar Abuse
16:00	Apr 6	Bosques del Lago	Familiar Abuse
16:25	May 19	Centro Urbano	Familiar Abuse
13:35	Mar 28	Hacienda del Parque	Local Shop Robbery
13:16	Jun 15	Infonavit Norte	Local Shop Robbery

Table 1. Fragment of report from the Attorney's Office.

Quadrant Num.	Time period
1	00:01:03:00 hrs.
2	03:01:06:00 hrs.
3	06:01:09:00 hrs.
4	09:01:12:00 hrs.
5	12:01:15:00 hrs.
6	15:01:18:00 hrs.
7	18:01:21:00 hrs.
8	21:01:00:00 hrs.

Table 2. Quadrants for times smoothing.

The crimes were clustered by kind in different classes, as previously mentioned in Section 2.1.2, for two different modes:

1. (1) Public roads and highways, (2) homes , (3) stores and shops.
2. (1) Robbery in all of its modalities, (2) homicide, (3) injury, (4) property damage.

The original data are now represented in Table 3, accordingly to Mode 1, and in Table 4, accordingly to Mode 2. For the first mode, we used only 205 from 1535 records because the remaining records did not have enough information to distribute them amongst the classes selected.

For the second mode, we were able to use all records. We will use this kind of information as the input of the KORA-Ω, CR-Ω and CR-Ω+ algorithms. The results of the classification of both analyses are reported in the following sections.

Time quadrant	Date	Residential Zone	Pub. Road (C 1)	Home (C 2)	Shops (C 3)
Q8	Jan	Arcos del Alba	1	0	0
Q4	Apr	Atlanta	1	0	0
Q1	Jun	Bosques de la Hda.	0	1	0
Q6	Apr	Bosques del Lago	0	1	0
Q6	May	Centro Urbano	0	1	0
Q5	Mar	Hacienda del Parque	0	0	1
Q5	Jun	Infonavit Norte	0	0	1

Table 3. Pre-processed report used as input to the algorithms —Mode 1.

Time quadrant	Date	Residential Zone	Robbery (C 1)	Injury (C 2)	Homicide (C 3)	Property Damage (C 4)
Q8	Jan	Arcos del Alba	1	0	0	0
Q4	Apr	Atlanta	1	0	0	0
Q1	Jun	Bosques de la Hda.	0	1	0	0
Q6	Apr	Bosques del Lago	0	1	0	0
Q6	May	Centro Urbano	0	0	1	0
Q5	Mar	Hacienda del Parque	0	0	0	1
Q5	Jun	Infonavit Norte	0	0	0	1

Table 4. Pre-processed report used as input to the algorithms —Mode 2.

3.2 Results of Punctual Hotspot prediction

In this section we present the results of our two first experiments consisting on applying the previously presented algorithms following the two modes mentioned in Section 2.1.2

3.2.1 First Experiment: Location of Crime

Mode 1 has three classes:
(1) Public roads and highways, (2) homes , (3) stores and shops.

3.2.1.1 Using the KORA-Ω Algorithm

Learning Rate. We calculate the characteristic features and the complementary features of the sample by applying the KORA-Ω to a set of data with 160 patterns (~78% of the 205 records from the whole sample). These 160 records were spread in the following way: Public roads: 105, Home: 35, Stores: 20. The learning percentage of the algorithm for the known data is 88%.

Prediction Rate. We used a sample of 45 patterns (~22% of the 205 records from the whole sample) divided as follows: 30 patterns for crimes in public roads, 8 patterns for home crimes, and 7 patterns for shop crimes. The algorithm had an effectiveness of 66% for the test set.

3.2.1.2 Using the CR-Ω+ Algorithm

Learning Percentage. We use the same 160 patterns applied in the previous experiment. The learning percentage of the algorithm for the known data rose to 92.5%.

Prediction Rate. We used the 45 patterns for test from the previous experiment. The algorithm had a learning rate of 69% of for the real data test set. This means that it was possible to predict in more than two thirds of cases the place where crimes are likely to have a greater incidence.

3.2.2 Second Experiment: Crime families

One of the main disadvantages depicted in the first analysis is the low number of records that can be used for prediction, although this allowed predicting the place where crimes are more likely to be perpetrated. For this experiment, we grouped the sample data in the following classes: (1) robbery in all of its modalities, (2) homicide, (3) injury, and (4) property damage. This analysis allows using 1231 records of 1551. For the sake of considering the heterogeneous distribution of the data between different dates, we tested the algorithms against:

I. 150 events corresponding to the month of April of 2007
II. 123 events corresponding to the month of July of 2007
III. 321 events corresponding to the month of April of 2008

In contrast with the previous experiment, where the 1551 records from January 1st , 2007 to July 31st, 2007 were split in 78% for training and 22% for testing, in this experiment we used 100% of such data as training, and the additional patterns of I, II and III as different tests. The purpose of testing against different test sets is to examine the performance of the algorithm given the heterogeneity of the provided data. It can be seen, for example, that the newest data from April 2008 has the double the number of records compared to data from previous dates (April of 2007, and July of 2007). For variant II, the month of July was not considered in the training set.

3.2.2 1 Results

Table 5 to Table 8 show the results of applying algorithms CR-Ω+ and CR-Ω+M to the different test sets. We also explore limiting the number of features in a feature set to at least two, as in the example of Section 2.2.1, and applying no limitations (so that feature sets can be composed of only one feature). For all experiments, the empirically chosen values for beta were $\beta_1^+ = 3$, $\beta_1'^+ = 1$, $\beta_1^- = 3$ and $\beta_1'^- = 1$, $\beta_2^+ = 0$, $\beta_2^- = 3$ for CR-Ω+ and $\beta_2^- = 1$ for CR-Ω+M

	CR-Ω+		**CR-Ω+M**		**CR-Ω+**		**CR-Ω+M**	
	(2+ Features)		**(2+ Features)**		**(1+ Features)**		**(1+ Features)**	
	Train	Test	Train	Test	Train	Test	Train	Test
robbery	334	54	334	54	334	54	334	54
injuries	439	60	439	60	439	60	439	60
homicide	36	3	36	3	36	3	36	3
property damage	272	33	272	33	272	33	272	33
total of records	1081	150	1081	150	1081	150	1081	150
records classified correctly	846	33	848	35	834	36	833	38
records classified incorrectly	235	117	233	115	247	114	248	112
coverage	91 %	100%	92 %	100%	91 %	100%	92 %	100%
precision (correct/covered)	86.0%	22%	85.1%	23%	84.8%	24%	83.6%	24%
recall (correct/total)	**77.0%**	**22%**	**78.0%**	**23 %**	**77.0%**	**24%**	**77.0%**	**24%**

Table 5. (I) 150 patterns corresponding to the month of April of 2007.

	CR-Ω+		**CR-Ω+M**		**CR-Ω+**		**CR-Ω+M**	
	(2+ Features)		**(2+ Features)**		**(1+ Features)**		**(1+ Features)**	
	Train	Test	Train	Test	Train	Test	Train	Test
robbery	338	50	338	50	338	50	338	50
injuries	455	44	455	44	455	44	455	44
homicide	31	8	31	8	31	8	31	8
property damage	284	21	284	21	284	21	284	21
total of records	1108	123	1108	123	1108	123	1108	123
records classified correctly	872	28	851	37	855	36	851	37
records classified incorrectly	236	95	257	86	253	87	257	86
covered records	1000	123	1012	123	1000	123	1012	123
uncovered recs.	108	0	96	0	108	0	96	0
coverage	90 %	100%	91%	100%	90%	100%	91 %	100%
precision (correct/covered)	87.2%	23%	84%	30%	85.5%	29%	84 %	30%
recall (correct/total)	**79.0%**	**23%**	**77%**	**30%**	**77.0%**	**29%**	**77%**	**30%**

Table 6. (II) 123 patterns corresponding to the month of July of 2007.

	CR-Ω+ (2+ Features)		**CR-Ω+M (2+ Features)**		**CR-Ω+ (1+ Features)**		**CR-Ω+M (1+ Features)**	
	Train	Test	Train	Test	Train	Test	Train	Test
robbery	388	135	388	135	388	135	388	135
injuries	499	97	499	97	499	97	499	97
homicide	39	13	39	13	39	13	39	13
property damage	305	76	305	76	305	76	305	76
total of records	1231	321	1231	321	1231	321	1231	321
records classified correctly	956	74	968	69	956	74	954	76
records classified incorrectly	275	247	263	252	275	247	277	245
covered records	1121	321	1133	321	1121	321	1133	321
uncovered recs.	110	0	98	0	110	0	98	0
coverage	91%	100%	92%	100%	91 %	100%	92%	100%
precision (correct/covered)	86%	23%	85.4%	21%	86%	23%	85.5%	24%
recall (correct/total)	**79%**	**23%**	**79.0%**	**21%**	**79%**	**23%**	**77.0%**	**24%**

Table 7. (III) 321 patterns corresponding to the month of April of 2008.

Recall	**CR-Ω+ (2+ Features)**		**CR-Ω+M (2+ Features)**		**CR-Ω+ (1+ Features)**		**CR-Ω+M (1+ Features)**	
I. April 2007	77.0%	22%	78.0%	23%	77.0%	24%	77.0%	**24%**
II. July 2007	79.0%	23%	77.0%	**30%**	77.0%	29%	77.0%	**30%**
III. April 2008	79.0%	23%	79.0%	21%	79.0%	23%	77.0%	**24%**

Table 8. Comparison of recall measures.

The CR-Ω+classifier was tested by classifying patterns previously not contained in the supervision sample [3]. We compared results against those achieved using the standard KORA-Ω algorithm and obtained an improvement for the learning rate as well as for the test rate. The original KORA-Ω algorithm obtained 88% and 66% for the learning rate and the test rate, respectively, whereas the proposed *CR*-Ω+algorithm obtained 92.5 and 69%, respectively, when classifying data into the following classes: (1) public roads, highways, (2) homes, (3) stores and shops. This suggests that we are able to predict the kind of crime spatially and temporally for two out of three crimes. We evaluated with test data for crimes perpetrated from January 1st , 2007 to July 31st, 2007. Approximately 78% was used for training and approximately 22% for testing.

In our second analysis, we used the whole data set from January 1st, 2007 to July 31st for training, while we selected three different data sets for

testing: **(I)** 150 patterns corresponding to the month of April of 2007, **(II)** 123 patterns corresponding to the month of July of 2007 and **(III)** 321 patterns corresponding to the month of April of 2008. We used three different datasets to evaluate the homogeneity of the data. We obtained 77% recall in the learning rate (up to 87% in precision), and 30% recall in forecast. This suggests that we are able to predict punctually the kind of crime, given a spatio-temporal location, at least for one of each four crimes perpetrated.
We proposed two new algorithms, the CR-Ω+ and its variation CR-Ω+M, which perform better than the original algorithms. Particularly, it can be seen from Table 8 that the CR-Ω+M improves in general the forecast recall—for example, for test II, using 2+ features, it raises recall from 23% to 30%.

3.3 Tendency Analysis

In this section, we present our tests with a different dataset and compare against predictions by the Naïve Bayes classifier. Additionally, we present results using the Spatio-Temporal Root Mean Square of Errors (STRMSE), measure proposed by Ivaha *et al.* [15]. This measure consists of daily measurements of forecast errors and it is based on the root mean squared error divided by the number of days of the sample. Two or more models may be compared using STRMSE as a measure of how well they explain a given set of observations: the unbiased model with the smallest STRMSE is generally interpreted as best explaining the variability in the observations. STRMSE is calculated as shown in Eq. (3). n is the total number of days forecasted and m is the total number of samples.

$$STRMSE = \sqrt{\frac{1}{n}\sum_{i}^{m}\frac{\left(O_i - \hat{O}_i\right)^2}{m}} \qquad (3)$$

3.3.1 Details of the Forecast Method

To test the proposed forecasting algorithm, the Sacramento dataset was used. This dataset contains 152,812 registered crimes and was made available by the Sacramento, CA, Police Department[1] . All crimes were committed within 19 surveillance sectors (space-units), over a period of time from January 2004 to December 2008 (time-units). Crimes were organized into 16 crime-families as shown in Table 9.

By analyzing only records from the last five years (2004 to 2008), a forecast was calculated for time-unit January 2009, all registered crime-families and within all 19 surveillance sectors. The foretold number of crimes was then compared with the real-life police registers from that same space-time unit (2,219 crimes during January 2009).

Two partial forecasts of the number *(F)* of crimes belonging to the *(c)* crime family to be observed within the *(s)* space unit and during *(t)* time unit are used to approximate the trend observed in the reference dataset. When the trend is known, it is possible to forecast the next value to be observed by looking at the last observed values. To achieve this forecast, the basic tool to be used is a query to the dataset to count the number of crimes, from a specific family (c) and observed within specific space and time units (s,t), the result of such a query is represented by function $\#\left(c,s,t\right)$.
Specifically, the foretold number (F) of crimes, belonging to the (C) crime-family, to be observed within the (S) space-unit and during time-unit (t), is determined by

$$F\left(c,s,t\right) = \alpha\left[f(c,s,t)\right] + \beta\left[g(c,s,t)\right] \qquad (4)$$

[1] http://www.sacpd.org/crime/stats/reports/

Where:

$f(c,t,s)$ Partial forecast yield by the analysis of crimes, from the same crime-family, observed within the same space-unit, but during the repeating time-unit (usually month of year) from each and every year available in the reference dataset (see below for more detail).

$g(c,t,s)$ Another partial forecast yield by the analysis of crimes (same crime-family and space-unit) observed during the last 12 months in the reference dataset (see below for more detail).

α, β Weight values assigned to each one of the previous partial forecast analysis.

By using this function, both partial forecasts can be expressed as

$$f(c,s,t) = \#\big(c,s,(t-1)\big) + \frac{1}{K_1} \sum_{p=last\_year}^{first\_year} \#(c,s,t^{p}) - \#(c,s,t^{p-1}) \tag{5}$$

Where t^{p} represents the last year present in the reference dataset.

$$g(c,s,t) = \#\big(c,s,(t-1)\big) + \frac{1}{K_2} \sum_{p=1}^{11} \#\big(c,s,(t-p)\big) - \#\big(c,s,(t-p+1)\big) \tag{6}$$

Where K_1 and K_2 are the number of relevant trend changes detected within the reference time-unit in each analysis, respectively.

3.3.2 Results

Using the above mentioned methodology, all positive and negative characteristic space-time properties for each crime-family were found. Results are shown in Table 9 and Table 10.

Crime family	**Train**	**Test**	**Bayes Forecast**	**Our Forecast**	**Bayes**	**Our Forecast**
	No. of patterns	**No. of patterns**	**Forecasted crimes**	**Forecasted crimes**	**(*STRMSE*)**	**(*STRMSE*)**
Burglary Vehicle	28,826	417	1647	405	13.92	1.95
Robbery	451	11	37	20	0.40	0.46
Burglary Residence	18,735	328	1363	364	10.93	1.27
Burglary business	7,827	135	578	148	4.88	0.72
Vandalism <$400	5,229	108	760	159	6.65	0.79
Vandalism >$400	10,269	80	338	103	2.64	0.60
Petty Theft	6,583	449	1937	424	15.71	1.54
Grand Theft	27,934	93	407	143	3.35	0.62
Traffic accident	10,610	170	716	223	5.64	0.77
Sexual Crimes	3,439	51	241	78	2.06	0.49
Domestic Violence	11,614	152	731	212	6.01	0.81
Gang Activity	3,005	51	226	33	2.10	0.40
Illegal Weapons	7,233	105	439	110	3.56	0.66
Homicide	348	2	15	7	0.24	0.14
Battery Civilians	5,695	67	372	63	3.18	0.47
Total	147,798	2,219	9,807	2,492	7.05	**0.90**

Table 9. Training from 1/1/2004 to 31/12/2008, Test month January 2009.

Crime family	Train	Test	Bayes Forecast	Our Forecast	Bayes	Our Forecast
	No. of patterns	No. of patterns	Forecasted crimes	Forecasted crimes	(*STRMSE*)	(*STRMSE*)
Burglary Vehicle	28,826	415	2005	388	17.75	2.16
Robbery	451	5	28	14	0.32	0.23
Burglary Residence	18,735	294	1394	349	11.40	1.77
Burglary business	7,827	89	670	142	6.41	0.85
Vandalism <$400	5,229	152	843	154	6.93	0.70
Vandalism >$400	10,269	61	385	90	3.24	0.51
Petty Theft	6,583	403	2172	401	18.52	1.40
Grand Theft	27,934	86	475	127	4.12	0.70
Traffic accident	10,610	159	928	218	8.25	0.84
Sexual Crimes	3,439	39	263	78	2.33	0.58
Domestic Violence	11,614	159	847	204	6.98	0.76
Gang Activity	3,005	60	238	44	2.24	0.40
Illegal Weapons	7,233	88	550	102	5.00	0.51
Homicide	348	1	21	13	0.26	0.21
Battery Civilians	5,695	57	415	64	3.73	0.36
Total	147,798	2,068	11,234	2,388	8.45	**0.97**

Table 10. Training from 1/1/2004 to 31/12/2008, Test month February 2009.

Table 11 is presented as a reference to the usage of the STMRSE measure. No direct comparison is intended as Ivaha *et al*. [15] worked with a different dataset, with the city of Cardiff, UK and in different time, though Eq.(3) normalizes the time-span by including the number of days into account.

	NFM	OLS-NI	OLS-PC	Ours
STMRSE	1.5745	1.1309	1.1313	0.97

Table 11. Rough comparison with the system presented by Ivaha *et al*.

These results show that the proposed method has very high effectiveness, with an STRMSE below 1.0 forecasting all space-units, during January 2009 (with a total of 2,219 crimes). This means that, in average, the proposed method only fails in less than five occurrences of each crime-family. Such precision is fairly acceptable for automated crime-analysis systems and might constitute a useful tool for planning preventive police operations.

4. Conclusions and Future Work

There are two main advantages of our model: first, the inductive definition of each class constructed by the learning process comprises by itself valuable information for describing the criminal scenarios under study, in addition, the flexibility of the β thresholds allow us to determine the level of precision we want in the inductive description of each class. Despite the fact that the CR-Ω+algorithm was not originally designed for dealing with time series data, constructing the supervision sample in strict chronologic order yields the same results. This means that the model will forecast new criminal scenarios very similar to those previously registered in the same space-time frame.

We performed two analyses: punctual prediction and tendency analysis, which show that it is possible to predict punctually one of four crimes to be perpetrated (crime family, in a specific space and time), and 66% of prediction of the place of crime, despite of the noise of the dataset. The tendency analysis yielded an STRMSE (Spatio-Temporal RMSE) of less than 1.0.

As future work, there are several paths to explore in this project. First, it is necessary to incorporate other information sources available. Second, it is of the utmost importance to calculate the optimal thresholds for the learning process. A statistical analysis of the data included in the supervision sample would make the task easier.

Finally, it is important to remark that the forecasting model herein exposed opens a very important research area for the logical-combinatorial classification methods [20, 10], one where, previously, only the statistical analysis has been used [8, 28].

References

[1] Amir, M., *Patterns in forcible rape*. Chicago: University of Chicago Press, 1971.

[2] Baldwin, J., Bottoms, A. *The urban criminal: A study in Sheffield*. London: Tavistock Publications, 1976.

[3] Baskakova, L.V., Y. I. Zuravlëv. Recognition Algorithm Models with Representative Sets and Supporting Sets Systems (in Russian), *Zh. Vichislitielnoi Matematiki i Matematicheskoi Fiziki*. Tom 21, No.5. URSS, 1981.

[4] Block, C., STAC hot-spot areas: A statistical tool for law enforcement decisions. In Block, C. R., Dabdoub, M., & Fregly, S. (Eds.), *Crime analysis through computer mapping.* Washington, DC: Police Executive Research Forum, p. 15−32, 1995.

[5] Bongard, M.N., et al, Solving geological problems using recognition programs. *Journal of Sovietic Geology,* No.6, 1963.

[6] Capone, D., Nichols, W., Urban structure and criminal mobility. *American Behavioral Scientist*, 20, 199–213, 1976.

[7] Carrasco-Ochoa, J.A., *Representative-Sets based Classifiers*, Masters Thesis, CINVESTAV-IPN, Mexico, 1994.

[8] Cheremesina, N. E., J. Ruiz Shulcloper, Cuestiones metodológicas de la aplicación de modelos matemáticos de Reconocimiento de Patrones en zonas del conocimiento poco formalizadas, *Revista Ciencias Matemáticas*, XIII(2), La Habana, Cuba, 1992.

[9] Cressie, N. A. C., *Statistics for spatial data*. Wiley, New York, 1993.

[10] De la Vega Doria, L., *Extension to the fuzzy case of the KORA-3 algorithm (in Spanish).* Masters Thesis, CINVESTAV-IPN, Mexico, 1994.

[11] De-la-Vega Doria, L., J. Ruiz Shulcloper, Carrasco Ochoa. Fuzzy KORA-Ω Algorithm. *Procedings of the 6th*

European Congress on Inteligent Techniques and Soft Computing, EUFIT, Aachen, Germany, 1998.

[12] Diukova, E.V., On a parametric model of KORA based recognition algorithms (in Russian), *Soovshenia po prikladmoi matematiki,* Russia, 1988.

[13] Goldfarb, L. A new approach in Pattern Recognition, *Progress in Machine Intelligence & Pattern Recognition,* Ed. L. Kanal, A. Rosenfeld. vol. II, 1985.

[14] Hirschfield, A., K. Bowers, *Mapping and Analysing Crime Data: Lessons from research and practice*, Taylor & Francis Inc., 2001.

[15] Ivaha, C., Al-Madfai, H., Higgs, G., Ware, A. and Corcoran, J., The simple spatial disaggregation approach to spatio-temporal crime forecasting. *International Journal of Innovative Computing Information And Control,* 3 3: 509–523, 2007

[16] Jefferis, E., A multi-method exploration of crime hot spots. *Presentation at the Annual Meeting of the Academy of Criminal Justice Sciences, Albuquerque, NM, March* 10–14, 1998.

[17] LeBeau, J. L., The journey to rape: Geographic distance and the rapist's methods of approaching the victim. *Journal of Police Science and Administration*, 15, 129–136, 1987.

[18] Leipnik, M., D.P. Albert, *GIS in Law Enforcement: Implementation issues and case studies*, Taylor & Francis Inc., 2003.

[19] Levine, N., "Hot Spot" analysis using CrimeStat kernel density interpolation. *Presentation at the Annual Meeting of the Academy of Criminal Justice Sciences, Albuquerque, NM, March* 10 –14, 1998.

[20] López-Reyes, N. Ruiz Shulcloper, *et al., Un sistema para el pronóstico a corto plazo de tormentas ionosféricas,* Reporte de investigación ICIMAF-ACC, 76, La Habana, CUBA, 1981.

[21] Martínez Trinidad, J.F., A. Guzmán Arenas, The Logical-Combinatorial approach to Pattern Recognition, an overview through selected works, *Pattern Recognition* 34:4, 2001.

[22] Mena, J., *Investigative Data Mining for Security and Criminal Detection*, Butterworth-Heinemann, 2003.

[23] Molumby, T., Patterns of crime in a university housing project. *American Behavioral Scientist*, 20, 247–259, 1976.

[24] Newman, O., *Defensible space: Crime prevention through urban design*. New York, Macmillan, 1973.

[25] Repetto, T. A., *Residential crime*. Cambridge, MA, Ballinger, 1972.

[26] Rossmo, D. K., Target patterns of serial murders: A methodological model. *American Journal of Criminal Justice,* 17(2), 1–21, 1993.

[27] Rossmo, D. K., Targeting victims: Serial killers and the urban environment. In O'Reilly-Flemming, T. (Ed.), *Serial and mass murder: Theory, research, and policy*. Toronto: Canadian Scholars Press, 1996.

[28] Ruiz Shulcloper J., Tutorial course: Classification with mixed and incomplete data (in Spanish). *VII Iberoamerican Conference on Pattern Recognition,* México, 2002.

[29] Ruiz Shulcloper, J., A. Guzmán Arenas, J. F. Martínez Trinidad, *Enfoque Lógico-Combinatorio al Reconocimiento de Patrones I: Selección de variables y Clasificación Supervisada.* CIC-IPN. Colección de Ciencia de la Computación, 1999.

[30] Scarr, H. A., *Patterns in burglary*. 2nd ed., Washington, DC: U.S. Department of Justice, 1973.

Authors´ Biography

Salvador GODOY-CALDERÓN

He graduated in computer engineering in 1992 at ITAM-Mexico. He received his master in sciences degree in 1994 from CINVESTAV and got his doctor's degree in computer sciences in 2006 from Center for Computing Research of the Instituto Politécnico Nacional (CIC, IPN), Mexico. His research interests are pattern recognition, testor theory, logic and formal systems. He has led several projects related to applied criminology in Mexico. Currently he is developing a pattern recognition suite based on the Combinatorial Logic approach at the Center of Computing Research.

Victor M.MARTÍNEZ-HÉRNANDEZ

Captain-Major of the Mexican National Defense Army. He graduated from Master of Sciences in 2009 from the Center of Computing Research of the National Polytechnic Institute (CIC, IPN), Mexico. His research interests are Data Mining and Pattern Recognition for crime analysis and prevention.

Marco A. MORENO- ARMENDARÍZ

Dr. Moreno-Armendáriz received the B.S. degree from La Salle University, Mexico in 1998 and the M.S. and Ph.D. degrees, both in Automatic Control, from CINVESTAV-IPN, México, in 1999 and 2003, respectively. From 2001 to 2006 he was researcher at the Engineering School in La Salle University, Mexico. In April of 2006, he joined the Center for Research in Computing of the National Polytechnic Institute, Mexico, where he is currently the head of the Pattern Recognition Laboratory. His research interests include Neural Networks for identification and control, Computer Vision, Robotics, Pattern Recognition and the implementation in FPGAs of related algorithms.

Hiram CALVO

He obtained his master’s degree in computer science in 2002 from the Universidad Nacional Autónoma de México (UNAM), with a thesis on mathematical modeling, and his Ph.D. degree in computer science (with honors) in 2006 from the Computing Research Center (CIC) of the Instituto Politécnico Nacional (IPN), Mexico, with a thesis on natural language processing. Since 2006, he has been a lecturer at the Computing Research Center (CIC) of the Instituto Politécnico Nacional (IPN). He was awarded with the National Lázaro Cárdenas Prize in 2006 as the best Ph. D. student of IPN in the area of physics and mathematics. Currently, he is a visiting researcher at the laboratory of Computational Linguistics, at the Nara Institute of Science and Technology, Japan.