

Modelos Matemáticos del Crimen

Alvaro J. Riascos Villegas
Universidad de los Andes y Quantil

Mayo 8 de 2018

Contenido

- 1 Modelos Matemáticos del Crimen
 - Validación
 - Aplicativo

- 2 Trabajo en Proceso

Contenido

- 1 Modelos Matemáticos del Crimen
 - Validación
 - Aplicativo
- 2 Trabajo en Proceso

Introducción

- Esta literatura a comenzado a influenciar la administración de los recursos policiales de los principales centros urbanos: Los Ángeles CA, Atlanta GA, Chicago IL, New York NW, Alhambra CA, San Francisco CA, Modesto CA, Santa Cruz, CA.
- Se diseñó una metodología para comparar diferentes modelos de predicción del crimen en Bogotá.
- Esta se extiende de forma natural a otros centros urbanos con la información adecuada (Cali, Medellín).
- Datos utilizados: 329,793 crímenes ocurridos en Bogotá entre 2004 y 2014 (geolocalizados, con fecha y hora).
- Se compararon modelos de Puntos, Elipses, KDE y varias versiones de modelos espacio-temporales.

Introducción

- Esta literatura a comenzado a influenciar la administración de los recursos policiales de los principales centros urbanos: Los Ángeles CA, Atlanta GA, Chicago IL, New York NW, Alhambra CA, San Francisco CA, Modesto CA, Santa Cruz, CA.
- Se diseñó una metodología para comparar diferentes modelos de predicción del crimen en Bogotá.
- Esta se extiende de forma natural a otros centros urbanos con la información adecuada (Cali, Medellín).
- Datos utilizados: 329,793 crímenes ocurridos en Bogotá entre 2004 y 2014 (geolocalizados, con fecha y hora).
- Se compararon modelos de Puntos, Elipses, KDE y varias versiones de modelos espacio-temporales.

Introducción

- Esta literatura a comenzado a influenciar la administración de los recursos policiales de los principales centros urbanos: Los Ángeles CA, Atlanta GA, Chicago IL, New York NW, Alhambra CA, San Francisco CA, Modesto CA, Santa Cruz, CA.
- Se diseñó una metodología para comparar diferentes modelos de predicción del crimen en Bogotá.
- Esta se extiende de forma natural a otros centros urbanos con la información adecuada (Cali, Medellín).
- Datos utilizados: 329,793 crímenes ocurridos en Bogotá entre 2004 y 2014 (geolocalizados, con fecha y hora).
- Se compararon modelos de Puntos, Elipses, KDE y varias versiones de modelos espacio-temporales.

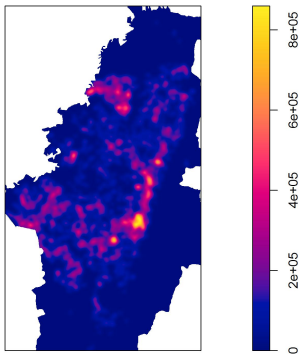
Introducción

- Esta literatura a comenzado a influenciar la administración de los recursos policiales de los principales centros urbanos: Los Ángeles CA, Atlanta GA, Chicago IL, New York NW, Alhambra CA, San Francisco CA, Modesto CA, Santa Cruz, CA.
- Se diseñó una metodología para comparar diferentes modelos de predicción del crimen en Bogotá.
- Esta se extiende de forma natural a otros centros urbanos con la información adecuada (Cali, Medellín).
- Datos utilizados: 329,793 crímenes ocurridos en Bogotá entre 2004 y 2014 (geolocalizados, con fecha y hora).
- Se compararon modelos de Puntos, Elipses, KDE y varias versiones de modelos espacio-temporales.

Introducción

- Esta literatura a comenzado a influenciar la administración de los recursos policiales de los principales centros urbanos: Los Ángeles CA, Atlanta GA, Chicago IL, New York NW, Alhambra CA, San Francisco CA, Modesto CA, Santa Cruz, CA.
- Se diseñó una metodología para comparar diferentes modelos de predicción del crimen en Bogotá.
- Esta se extiende de forma natural a otros centros urbanos con la información adecuada (Cali, Medellín).
- Datos utilizados: 329,793 crímenes ocurridos en Bogotá entre 2004 y 2014 (geolocalizados, con fecha y hora).
- Se compararon modelos de Puntos, Elipses, KDE y varias versiones de modelos espacio-temporales.

Heat map for a 1-year period of crime



- Este es un modelo basado en la clasificación de los eventos como antecedentes y réplicas.
- Es el estado del arte en modelos de predicción del crimen.

Modelo Espacio - Temporal: Motivación

Mohler et al.: Self-Exciting Point Process Modeling of Crime

101

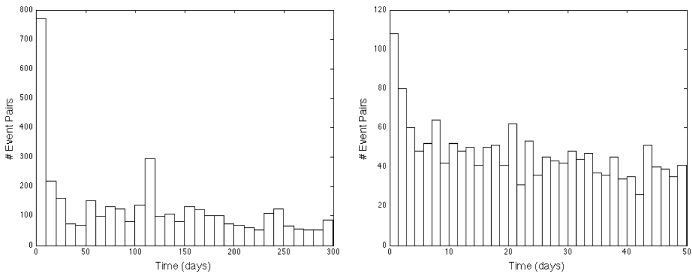


Figure 1. On the left, histogram of times (less than 300 days) between Southern California earthquake events of magnitude 3.0 or greater separated by 110 kilometers or less. On the right, histogram of times (less than 50 days) between burglary events separated by 200 meters or less.

Modelo Espacio - Temporal: Motivación

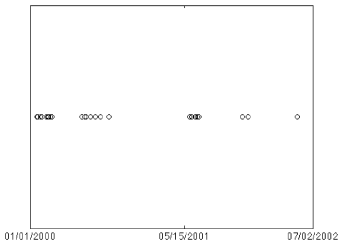


Figure 2. Times of violent crimes between two rivalry gangs in Los Angeles.

- Se considera un modelo de la intensidad espacio temporal del crimen de la forma:

$$\lambda(t, x, y) = \mu(t, x, y) + \sum_{k:t_k < t} g(t - t_k, x - x_k, y - y_k) \quad (1)$$

Validación

- Evaluar la capacidad predictiva del modelo de predicción de crimen propuesto.
- En particular, se compara el poder predictivo del modelo al ser entrenado con diferentes horizontes de tiempo.

- Se cuenta con datos de crímenes en Bogotá del 16 de abril al 30 de junio de 2017: 16.402 datos.
- Para esta validación se usaron datos de la localidad de Santa Fé, en donde se aplicará el piloto, delimitada por latitudes entre [4,571, 4,629] y longitudes $> -74,091$.

- Se cuenta con datos de crímenes en Bogotá del 16 de abril al 30 de junio de 2017: 16.402 datos.
- Para esta validación se usaron datos de la localidad de Santa Fé, en donde se aplicará el piloto, delimitada por latitudes entre $[4,571, 4,629]$ y longitudes $> -74,091$.

- 1.676 ($\approx 10\%$) crímenes en la localidad de Santa Fé en el período tratado.
- Se entrenó el modelo de crimen con datos entre 1 y 7 semanas y se validó con las 3 semanas posteriores al entrenamiento, en todos los casos, del 10 al 30 de junio de 2017 (407 crímenes).

- 1.676 ($\approx 10\%$) crímenes en la localidad de Santa Fé en el período tratado.
- Se entrenó el modelo de crimen con datos entre 1 y 7 semanas y se validó con las 3 semanas posteriores al entrenamiento, en todos los casos, del 10 al 30 de junio de 2017 (407 crímenes).

- Utilizamos el *Precision Accuracy Index*

$$\text{PAI} = \frac{\text{Hit Rate}}{\text{Percentage of Area}}$$
$$\text{Hit Rate} = \frac{\text{Crimes predicted in Hotspots}}{\text{Total Crimes}}$$
$$\text{Percentage Area} = \frac{\text{Area of Hotspots}}{\text{Total Area}}$$

- Sin embargo en modelos muy granulares no es una buena medida.

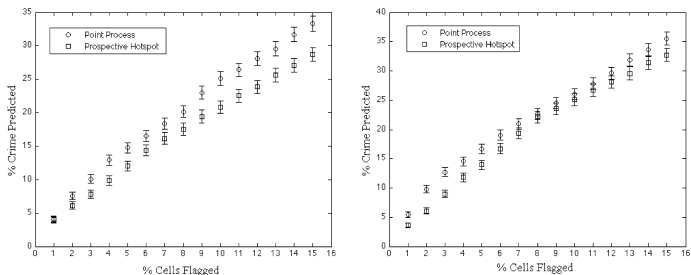


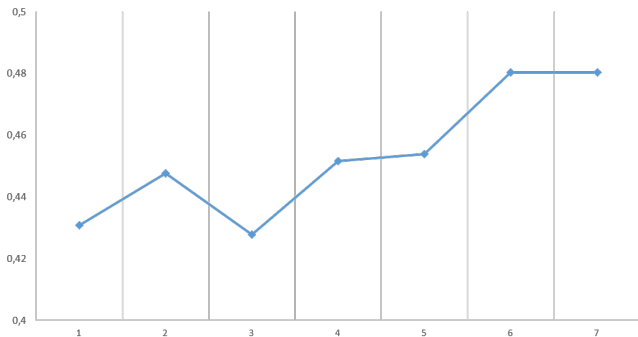
Figure 6. Forecasting strategy comparison. Average daily percentage of crimes predicted plotted against percentage of cells flagged for 2005 burglary using 200 m by 200 m cells. Error bars correspond to the standard error. Prospective hotspot cutoff parameters are 400 meters and 8 weeks (left) and optimal parameters (right) are 200 meters and 39 weeks. Spatial background intensity $\mu(x, y)$ smoothing bandwidth for the point process is 300 meters (left) selected by cross validation and 130 meters (right) selected to optimize the number of crimes predicted.

- Se divide Bogotá ($1,547\text{km}^2$) en 10,946 celdas de $\approx 145\text{m}^2$ cada una; 1,019 celdas en la localidad de Santa Fé.
- Se entrena el modelo con el correspondiente número de semanas y se predicen los puntos calientes para cada turno de 8 horas, definidos como el 10% de las celdas con mayor probabilidad de crimen.
- Se investigan cuántos crímenes de los datos de validación ocurrieron en los puntos calientes predichos por el modelo.

- Se divide Bogotá ($1,547\text{km}^2$) en 10,946 celdas de $\approx 145\text{m}^2$ cada una; 1,019 celdas en la localidad de Santa Fé.
- Se entrena el modelo con el correspondiente número de semanas y se predicen los puntos calientes para cada turno de 8 horas, definidos como el 10% de las celdas con mayor probabilidad de crimen.
- Se investigan cuántos crímenes de los datos de validación ocurrieron en los puntos calientes predichos por el modelo.

- Se divide Bogotá ($1,547\text{km}^2$) en 10,946 celdas de $\approx 145\text{m}^2$ cada una; 1,019 celdas en la localidad de Santa Fé.
- Se entrena el modelo con el correspondiente número de semanas y se predicen los puntos calientes para cada turno de 8 horas, definidos como el 10% de las celdas con mayor probabilidad de crimen.
- Se investigan cuántos crímenes de los datos de validación ocurrieron en los puntos calientes predichos por el modelo.

HIT RATE PROMEDIO POR SEMANAS DE ENTRENAMIENTO



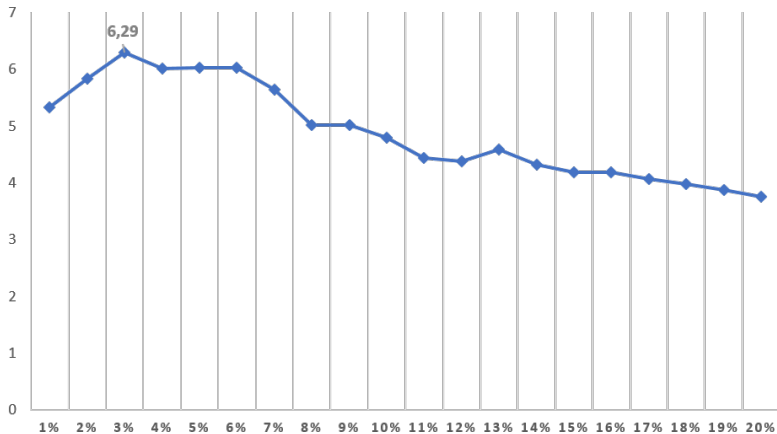
- Hit Rate con 7 semanas de entrenamiento y 10% de cobertura de puntos calientes:

Predicción	bw fijo	bw variable	KDE
Semana 1	0,44	0,57	0,42
Semana 2	0,46	0,59	0,44
Semana 3	0,54	0,62	0,53
Promedio	0,48	0,59	0,46

- Hit Rate con 7 semanas de entrenamiento y 10% de cobertura de puntos calientes:

Predicción	bw fijo	bw variable	KDE
Semana 1	0,44	0,57	0,42
Semana 2	0,46	0,59	0,44
Semana 3	0,54	0,62	0,53
Promedio	0,48	0,59	0,46

PAI PROMEDIO SEGÚN PORCENTAJE DE COBERTURA



~/R/Code/codigo_crimen/App - Shiny
http://127.0.0.1:5084 Open in Browser Publish

Predicción de crimen

quartil matemáticas aplicadas

Horas: 10

Día:
 lunes
 martes
 miércoles
 jueves
 viernes
 sábado
 domingo

Código del cuadrante:
E01-13

Ejecutar

Mapa de calor Patillaje

Powered by Leaflet — © OpenStreetMap contributors, CC-BY-SA

Patrullaje por segmentos

Predicción de crimen

quantil

matemáticas aplicadas

Horas:



Día:

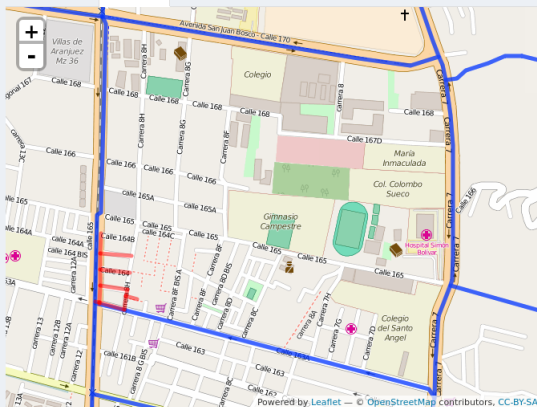
- lunes
- martes
- miércoles
- jueves
- viernes
- sabado
- domingo

Código del cuadrante:

E01-12

Ejecutar

Mapa de calor | Patrullaje



Patrullaje por segmentos

Contenido

- 1 Modelos Matemáticos del Crimen
 - Validación
 - Aplicativo

- 2 Trabajo en Proceso

- Salud pública:
 - 1 Grupos de enfermedades usando análisis supervisado: Aprendizaje de máquinas, distancia entre particiones y optimización usando el algoritmo Metropolis Hasting.
 - 2 Grupos de enfermedades usando análisis no supervisado: LDA.
- Predicción del crimen (en conjunto con el Profesor Francisco Gomez).
 - 1 Teoría de las ventanas rotas, imágenes y aprendizaje profundo.
 - 2 Viajes de Levy.
 - 3 Kernel aprendizaje en variedades.

- Salud pública:
 - 1 Grupos de enfermedades usando análisis supervisado: Aprendizaje de máquinas, distancia entre particiones y optimización usando el algoritmo Metropolis Hasting.
 - 2 Grupos de enfermedades usando análisis no supervisado: LDA.
- Predicción del crimen (en conjunto con el Profesor Francisco Gomez).
 - 1 Teoría de las ventanas rotas, imágenes y aprendizaje profundo.
 - 2 Viajes de Levy.
 - 3 Kernel aprendizaje en variedades.

- Salud pública:
 - 1 Grupos de enfermedades usando análisis supervisado: Aprendizaje de máquinas, distancia entre particiones y optimización usando el algoritmo Metropolis Hasting.
 - 2 Grupos de enfermedades usando análisis no supervisado: LDA.
- Predicción del crimen (en conjunto con el Profesor Francisco Gomez).
 - 1 Teoría de las ventanas rotas, imágenes y aprendizaje profundo.
 - 2 Viajes de Levy.
 - 3 Kernel aprendizaje en variedades.

- Salud pública:
 - 1 Grupos de enfermedades usando análisis supervisado: Aprendizaje de máquinas, distancia entre particiones y optimización usando el algoritmo Metropolis Hasting.
 - 2 Grupos de enfermedades usando análisis no supervisado: LDA.
- Predicción del crimen (en conjunto con el Profesor Francisco Gomez).
 - 1 Teoría de las ventanas rotas, imágenes y aprendizaje profundo.
 - 2 Viajes de Levy.
 - 3 Kernel aprendizaje en variedades.