

Early Warning System for Temporary Crime Hot Spots

Carlos Díaz

Quantil

carlos.diaz@quantil.com.co

January 14, 2016

- 1 Ojetivos del paper
- 2 Revisión de la literatura
- 3 Diseño experimental
- 4 Algunos resultados
- 5 Estimación por kernel

Objetivos del paper

- Poner a prueba un sistema de prevención de crímenes basado en el despliegue/detención de recursos policiales en hot spots temporales (y se mantiene constante en hot spots crónicos).
- Investigar el potencial del método implementado para prevenir la aparición y el desarrollo de hot spots temporales y crónicos.

- Los hot spots temporales
- El mapeo de hot spots
- Métodos de detección y predicción del crimen

Los hot spots temporales

Se definen como zonas cuyas densidades de crimen (crimen por área por unidad de tiempo) son comparables con las de los hot spots crónicos, pero normalmente persisten por periodos de tiempo menores a un año.

Normalmente estos hot spots aparecen, persisten por un tiempo, se extinguen, y a veces pueden reaparecer luego de cierto tiempo.

El sistema de prevención implementado en el paper busca prevenir la aparición y/o persistencia de hot spots temporales.

Existen varias posibilidades para mapear hot spots (Chainey, 2008):

- Mapeo por puntos.
- Elipses espaciales.
- Mapas temáticos por unidades administrativas.
- Mapeo temático por grilla.
- Estimación de densidades por kernels.

La efectividad del mapeo

La manera como se compara cuál de los métodos de mapeo es mejor se hace usando el PAI (*prediction accuracy index*)

$$PAI = \frac{Hit\ rate}{Porcentaje\ de\ area} \quad (1)$$

Donde

$$Hit\ rate = \frac{crimenes\ predichos\ (acertados)}{crimenes\ totales} \quad (2)$$

$$Porcentaje\ de\ area = \frac{area\ de\ prediccion}{area\ total} \quad (3)$$

Detección y predicción del crimen

La literatura sobre como detectar/predecir el comienzo de una temporada de crimen es bastante limitada.

La mayoría de los modelos que se ven en la literatura son extrapolativos. Se toma una serie temporal de datos sobre crímenes pasados y se trata de extrapolar para poder detectar/predecir crímenes futuros.

Algunos modelos "causales" se pueden encontrar en la literatura. Estos modelos son multivariados, y consideran independientes a las variables. Dentro de ellos se destacan algunos por su aparente éxito (Cohen 2007; Gorr 2009). Estos toman como punto de referencia a los eventos/crímenes leves que vienen en aumento y los consideran causales de crímenes más graves. Se hace testeo de hipótesis (estadística) basado en extrapolación.

- Datos: Base de datos con 329.793 crímenes entre marzo de 2004 - marzo de 2014, organizados segun latitud, longitud, año, mes, día, día de la semana, hora exacta y tipo (14 tipos).
- Los datos de crímenes se organizan de manera **mensual**.
- Datos experimento: 41.961 crímenes entre enero de 2011 - febrero de 2012.
- Por facilidad, se organizan los datos en una grilla (raster) que cubre la ciudad. Para el lado de cada cuadro de la grilla se manejan como opciones 150, 225 y 300 metros (que es la medida aproximada para 4, 9 y 16 cuadras de area, respectivamente).

Regla de detección

Para determinar si una zona es un hot spot se escoge un umbral de crímenes para detectar en primer lugar. A continuación se presenta una tabla con la frecuencia de crímenes que detecta cada una de las opciones para el tamaño de las celdas:

Tamaño de la celda (m)	1 crimen	2 crímenes	textbfmayor a 2
150	1826	682	425
225	1342	838	753
300	990	832	1111

Table: Posibles umbrales para detectar el inicio de un hot spot (enero 2011)

La información que más se ajusta a la realidad parece ser cuando el tamaño de la celda es de 150 metros. Más aún, se escogerá un umbral de 2 crímenes (2 o más representaría tan solo el 14,5% de los casos, lo cual es muy bajo).

Regla de predicción

Una vez que se hace la detección, se debe tener una regla para intentar predecir si una zona puede convertirse en un hot spot temporal. A continuación se muestra una tabla que muestra el número de celdas con 2, 3, 4 y 5 meses de crímenes a lo largo de cierto año (el raster cuenta con 39750 celdas):

Cantidad de meses	Número de celdas
2	1071
3	479
4	267
5	150

Table: Celdas que detectaron crímenes en 2, 3, 4 o 5 meses

Ahora, si usamos la regla de detección en el siguiente mes (después del año de observaciones), el número de celdas detectadas que reportan crímenes en al menos 2, 3, 4 o 5 meses en el año previo está dado así:

Cantidad de meses	Número de celdas
2	257
3	182
4	124
5	86

Table: Posibles umbrales para confirmar la aparición de un hot spot temporal

Algunos resultados

A continuación se muestra el número de celdas expuestas al trabajo de prevención cuando se elige cada uno de los umbrales en la regla de predicción y el porcentaje de crímenes expuestos a prevención en ese mes:

Umbral (meses)	Celdas con prevención	Porcentaje
2	167	30%
3	125	22,7%
4	90	16,3%
5	68	12,4%

Table: Posibles umbrales para confirmar la aparición de un hot spot temporal

El modelo de estimación por kernel toma los datos de crímenes y los considera como un procesos de puntos marcados. Se crea una función intensidad, que es la que se usa para determinar los umbrales. Tendríamos de nuevo:

- Regla de detección
- Regla de predicción

Comparación entre modelos: un ejemplo

A continuación se muestran tablas en las que se calculan los PAI para un caso de criminalidad a lo largo de cierto día en enero de 2012

Intervalo de tiempo	PAI
00:00 - 02:59	3.2
03:00 - 05:59	0
06:00 - 08:59	4.68
09:00 - 11:59	3.45
12:00 - 14:59	6.88
15:00 - 17:59	10.86

Table: índice PAI calculado para estimación por kernel

Comparación entre modelos: un ejemplo

Intervalo de tiempo	PAI
00:00 - 02:59	1.02
03:00 - 05:59	0
06:00 - 08:59	1.7
09:00 - 11:59	0.06
12:00 - 14:59	2.42
15:00 - 17:59	4.65

Table: índice PAI calculado para modelo de puntos



Chainey S, Tompson L, Uhlig S (2008)

The utility of hotspot mapping for predicting spatial patterns of crime.
Security J. 21, 4 – 28.



Cohen J, Gorr WL (2007)

Leading indicators and spatial interactions: a crime-forecasting model for proactive police deployment.
Geogr Anal 39, 105 – 127.



Gorr WL (2012)

Forecast accuracy measures for exception reporting using receiver operating characteristic curves.
Int J Forecast 25, 48 – 61.

¡GRACIAS!