

Muchas emergencias y aún más llamadas

Rafael Prieto Curiel

Centro de Atención a Emergencias y
Protección Ciudadana
de la Ciudad de México
rafaelprietocuriel@yahoo.com

Irving Simonin Wilmer

Jefe de Unidad Departamental,
Centro de Atención a Emergencias y
Protección Ciudadana de la Ciudad de México
isimoninw@gmail.com

Resumen

Optimizar la distribución y el despacho de los recursos es un tema prioritario en la atención a emergencias, por lo que identificar las posibles llamadas que provienen de un mismo evento resulta determinante para su correcta atención. Se analizará el problema tomando en cuenta las características de una llamada de emergencia y los datos disponibles a partir de los cuales se pueden relacionar distintos reportes y, mediante un modelo de regresión logística, se encontrarán criterios óptimos para relacionar dos reportes.

Introducción

En la Ciudad de México se atienden diariamente miles de emergencias, producto del día a día y del trabajo cotidiano y rutinario de una gran urbe. Brindar la mejor atención posible y lograr que llegue una ambulancia, un camión de bomberos o una patrulla, según lo requiere la emergencia, es responsabilidad de los diversos actores de gobierno y por ello es una prioridad de las instituciones detectar de manera oportuna el lugar en el que se suscitan las situaciones de riesgo para la población.

Para ese cometido, existen diversas maneras de aumentar la detección de emergencias, como cámaras de seguridad o botones de pánico,

aunque sin duda, la mejor de ellas es un único número de emergencias en el cual se puedan realizar reportes y solicitar auxilio. De la misma manera en la que funciona el número 911 en Estados Unidos o el 112 en los países de la Unión Europea, en México se pueden reportar emergencias y requerir auxilio marcando el teléfono 066.

En algunos incidentes de gran magnitud, como un incendio o una explosión, es muy común recibir más de una llamada (si no es que decenas o incluso cientos) alertando sobre la misma situación y resulta sumamente importante, con el objetivo de no asignar más recursos de los necesarios para su atención, identificar los reportes repetidos, pero incluso resulta más importante detectar los eventos que, aunque estén cerca y podrían parecer la misma emergencia, están aislados y se trata de dos distintas situaciones con dos requerimientos de recursos. Asumir de manera incorrecta que dos llamadas provienen de la misma emergencia puede resultar catastrófico para la atención de alguna de ellas.

Por lo anterior, es necesario que el despachador encargado de la emergencia use su criterio para determinar si dos llamadas están ligadas. El proceso para determinar si dos llamadas provienen de la misma emergencia se realiza mediante un sistema que, basado en un criterio simple—como la distancia entre los eventos— sugiere al operador eventos que pueden estar relacionados al incidente que está atendiendo y el operador utiliza su criterio y experiencia para decidir si se trata de la misma situación. Dada la enorme cantidad de llamadas es importante que las sugerencias realizadas al operador sean adecuadas y que sea la menor cantidad posible, pues revisar cada uno de los incidentes requiere tiempo del operador. Por ello, nos enfocaremos en los incidentes que son sugeridos al despachador a partir de una llamada inicial.

Nuestra información

Antes de intentar diseñar un modelo para identificar las posibles llamadas duplicadas, es importante entender el origen mismo de las variables disponibles, es decir, entender una llamada de emergencia y la información que de ella se desprende. Frecuentemente el usuario del número de emergencia comunica la información que tiene disponible a la vista, sin que necesariamente sea precisa o incluso cierta. Un ejemplo de ello es recibir el reporte de una persona que aprecia una columna de humo, sin que pueda identificar el domicilio exacto en el que ocurre el incendio; por ello ocasionalmente se reciben una serie de llamadas reportando muchos incendios, todos ellos relativamente cercanos. Otro ejemplo es el sonido de una explosión, que de la misma manera genera varias llamadas, algunas incluso reportando disparos o personas lesionadas.

Adicionalmente, existen ciertos incidentes que pueden ser grandes en cuanto a su área de impacto, como una explosión, o largos en cuanto a su duración, como un incendio forestal, y por ello, es frecuente atender llamadas que hacen referencia al mismo evento con varios minutos de diferencia entre ellas, por ejemplo, una inundación.

Es importante entender que los llamantes reportan lo que tienen a la vista y que todos proporcionan información que puede ser relevante para la atención del incidente.

Identificando llamadas relacionadas

El objetivo es identificar las llamadas que provienen del mismo incidente, por lo que se pueden cometer dos distintos tipos de error:

- 1) No ligar dos llamadas que provienen de la misma emergencia. Ese tipo de error sucede por lo general en incidentes que se pueden ubicar en espacio amplio o con movimiento —como una manifestación o una marcha— o que tienen una duración prolongada como una inundación.
- 2) El segundo tipo de error consiste en relacionar eventos que no están ligados. Asumir que, por estar relativamente cercanos tanto en espacio como en tiempo, dos llamadas corresponden a la misma emergencia. Esa situación es propensa a suceder en lugares y momentos con una alta densidad de llamadas, que por lo general son lugares con flujo considerable de personas.

Considerar únicamente un tipo de error e intentar minimizarlo tiene una obvia —pero ilógica— respuesta: ligar todas llamadas entre sí (para el caso del primer error) o no relacionar ninguna (para el segundo tipo de error). Por ello es importante lidiar con ambos simultáneamente y ponderarlos correctamente.

Definiendo el problema

De manera más formal, supongamos que tenemos una llamada inicial, c_0 , a la cual queremos buscar los posibles reportes ligados, es decir, se busca calcular la probabilidad de que cada incidente esté relacionado a c_0 y determinar un umbral de probabilidad a partir del cual, si se excede, asumiremos que las llamadas corresponden al mismo incidente y se atenderán como uno solo, y en otro caso se deberán de atender como eventos aislados y mandar dos recursos diferentes.

Debido a que la consulta de emparejamiento de llamadas se hace para cada llamada por separado, el problema se puede representar como uno

de clasificación: el conjunto de llamadas que están relacionadas a la llamada c_0 y las que no. Entonces, dejando fija c_0 , se busca definir un criterio apropiado para elegir la zona geográfica, un umbral de tiempo y el tipo de incidentes en la cual se encuentran las posibles llamadas ligadas a c_0 . Para lograr el objetivo, tomaremos en cuenta las siguientes variables para cada una de las llamadas c_i .

1. Distancia entre el lugar reportado por las llamadas c_i y c_0 .
2. Tiempo transcurrido entre las llamadas.
3. «Distancia» en tipo de incidente reportado.

La tercer variable se refiere a qué tan parecido es un incidente a otro o qué tan posible es que estén relacionados. El ejemplo con el que se comienza el documento sugiere que una persona lesionada puede estar relacionada con una explosión previamente reportada; en contraste, esa misma persona lesionada es más improbable que esté relacionada a un robo de vehículo. Debido a que se cuenta con una fuente de datos rica en información, la forma de calcular la «distancia» en tipo de incidente que se considera es un simple conteo de relaciones entre los tipos de incidentes (e.g. ¿Cuántas veces se han ligado explosiones con incendios en el pasado?).

Ahora se define y_i como la probabilidad de que alguna llamada c_i esté relacionada a c_0 . El propósito es estimar y_i a partir de los datos que se tienen disponibles. El método elegido es una regresión logística, la cual pertenece a la familia de modelos lineales generalizados, los cuales consisten en estimar una variable dependiente y a partir de un conjunto de p variables llamados regresores x_1, x_2, \dots, x_p , suponiendo una relación entre una combinación lineal de los regresores y una función de enlace g . O escrito de otra manera:

$$g(y) = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p.$$

En el caso de la regresión logística, la función de enlace es la llamada función *logit* (de ahí el nombre) con la cual podemos representar el modelo final:

$$g(y) = \text{logit}(y) = \log\left(\frac{y}{1-y}\right) = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p.$$

La imagen de la función inversa de la función logit

$$g^{-1}(x) = \frac{1}{1 + e^{-x}},$$

está contenida en el intervalo $(0, 1)$ y es ideal para, por ejemplo, fijar umbrales de decisión en problemas de clasificación binarios.

Para esa probabilidad, que es estimada a partir de las variables que se tienen disponibles de cada llamada, se definirá un *punto de corte* a

partir del cual se clasificarán dos llamadas como provenientes del mismo incidente.

Minimizando el área

Además de las llamadas que están relacionadas de manera «obvia», que son aquellas que están a muy poca distancia y muy poco tiempo entre ellas, existe una estrecha correlación entre esas variables, es decir, las llamadas relacionadas generalmente son reportadas en el mismo lugar, aunque con una considerable diferencia en el tiempo o son realizadas prácticamente al mismo instante, aunque pueden estar reportadas en lugares lejanos. Esa relación se puede apreciar en la figura 1 en la que se expresan la distancia y el tiempo entre dos llamadas relacionadas al mismo incidente. Por ello, el resultado esperado de este proyecto es definir una zona en la cual los incidentes que estén dentro se consideren relacionados con el incidente c_0 y las llamadas que estén fuera de ella, se traten como independientes. Los incidentes que no están relacionados a c_0 no presentan ninguna relación entre el espacio o el tiempo que son reportados, por lo que con el objetivo de minimizar el segundo tipo de error (ligar incidentes independientes), el criterio empleado es buscar que esa zona ocupe el menor *espacio* posible.

Distancia y tiempo entre llamadas relacionadas

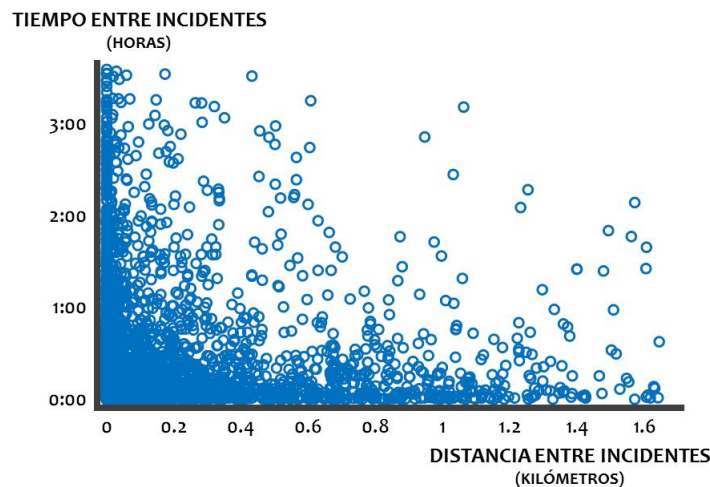


Figura 1. Esta gráfica muestra la distancia y la diferencia en tiempo de llamadas que sí están relacionadas entre sí. Cada punto representa una pareja de llamadas.

Entrenamiento

Se utilizó un conjunto de entrenamiento que consiste en un grupo de miles de llamadas que, al analizarlas de manera manual, se encontró que estaban relacionadas entre sí. Con el objetivo de minimizar el área de emparejamiento, se crearon, dentro de la región a analizar, una cantidad igual al número de llamadas emparejadas, distribuidas de manera aleatoria. Esos puntos representan «ruido», es decir, las confusiones en las que se podría caer al ligar dos llamadas independientes. Al enfrentar los puntos que efectivamente están relacionados entre sí con el ruido se obtiene un problema similar al de minimizar el área de emparejamiento: mientras más grande sea el área, más «ruido» tendrá dentro y se estarán cometiendo más errores de emparejamiento.

Para elegir el punto de corte PC que representa el umbral a partir del cual se determina que dos incidentes están ligados, es necesario determinar una función que penalice ambos tipos de error, o en nuestro caso, que considere el error u el área de la región de aceptación. Para ello se definió la siguiente función de eficiencia:

$$Ef(PC) = \frac{\text{Emparejamientos en } AC}{\text{Total emparejamientos}} - \frac{\text{Ruido en } AC}{\text{Total de elementos en el ruido}},$$

donde AC es el área crítica definida por los parámetros y el punto de corte. Con esto, para encontrar el punto de corte óptimo, debemos elegir aquél que maximice la eficiencia. En la figura 2 se puede apreciar un simple esquema sobre el efecto de elegir distintos puntos de corte, así como una simple idea para escoger el punto de corte óptimo. De esta manera, si tomamos PC cercano a cero, entonces prácticamente todos los eventos relacionados estarán dentro del área crítica, pero al mismo tiempo, casi todo el ruido también estará dentro. Es decir, tomamos un área muy grande. Si por el contrario, tomamos PC cercano a uno, entonces casi todo el ruido estará fuera, ya que el área será muy pequeña, pero al mismo tiempo casi nunca relacionaremos dos llamadas.

Una vez elegido el punto de corte óptimo, este define una curva que representa la frontera del AC en la cual dos llamadas serán consideradas como el mismo incidente y es necesario considerar que distintos puntos de corte definen distintas curvas y regiones de emparejamiento. Se realizaron pruebas con diferente ruido para evitar que casos extremos definan el AC y se tomó la curva promedio como la frontera de la región crítica, como se puede observar en la figura 3.

Efecto de elegir diferente Punto de Corte

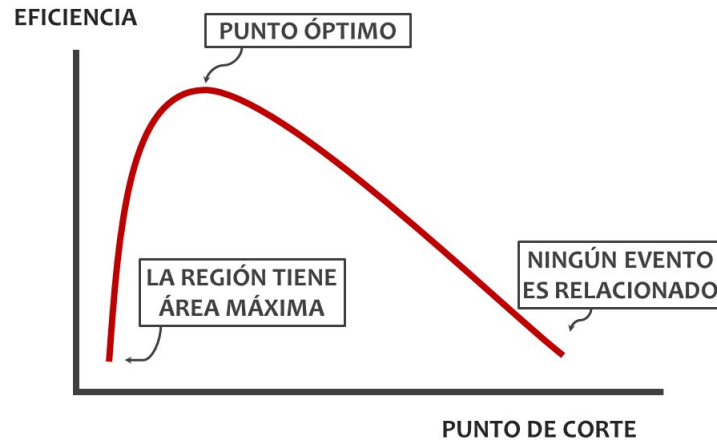


Figura 2. Gracias a que se considera el error de clasificación y el área crítica se obtiene un punto de corte óptimo para el problema de ligar incidentes.

Resultados de clasificación del modelo

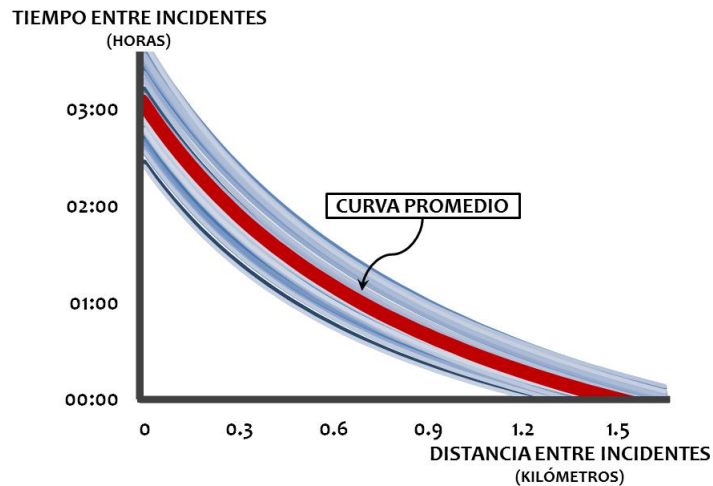


Figura 3. Representación de la curva definida por los puntos de corte para distintos soportes del ruido. Solo se ilustran las dimensiones de distancia y tiempo y el término cuadrático distancia \times tiempo. La región crítica para relacionar incidentes se encuentra debajo de la curva.

Resultados

Gracias al criterio que se obtiene a partir del modelo anterior, el resultado de la clasificación tiene dos grandes impactos. El primero es que

en solo una pequeña parte de las llamadas se realizará una sugerencia de otros eventos que posiblemente estén relacionados; es decir, hay una reducción considerable de opciones que se traduce en un ahorro del tiempo de operación. El operador tendrá que revisar muchos menos incidentes de los que se hubieran presentado anteriormente. Además, de las opciones presentadas, 83 % son llamadas que efectivamente provienen del mismo evento, es decir, las sugerencias para ligar son altamente precisas y se evita relacionar incidentes que son independientes. Esto ayuda a que el despachador tenga una menor posibilidad de error en una menor cantidad de tiempo, dando mejor atención a la ciudadanía. En la figura 4 se reflejan los incidentes que son correcta e incorrectamente clasificados con el modelo desarrollado.

Llamadas relacionadas y la predicción del modelo final

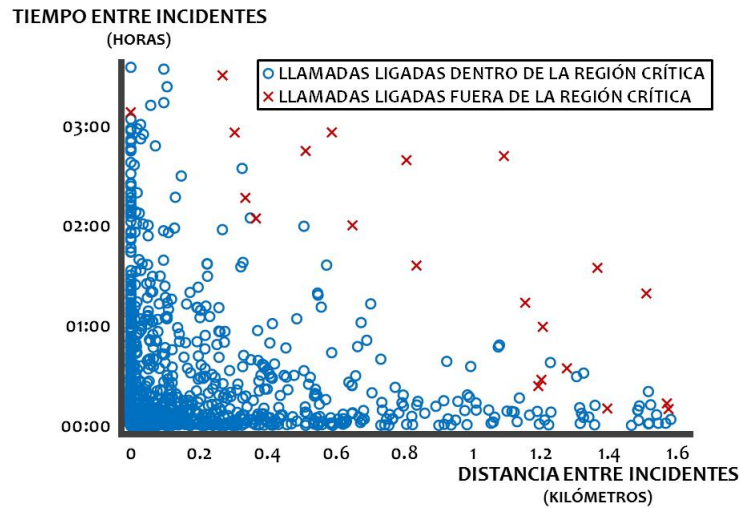


Figura 4. Clasificación de parejas relacionadas. Las 4 dimensiones finales del modelo están consideradas: distancia, tiempo, tipo y distancia \times tiempo. Los anillos son clasificados como relacionados y las \times representan los emparejamientos que el modelo deja fuera.

Presentar menos sugerencias conlleva un costo, pues algunos incidentes ligados caen fuera de la región que se sugiere al operador. Cerca de una tercera parte de los incidentes que efectivamente están ligados no son sugeridos al operador, aunque usualmente ellos representan eventos «especiales», pues son reportados lejos del evento inicial y con un tiempo considerable de desfase, es decir, son eventos espacialmente grandes y largos en duración, como un incendio forestal. En ese caso, dada la relevancia y magnitud que pueden tener, no representa una gran pérdida que sean despachados varios recursos al lugar.

Conclusiones

En el ámbito de atención a emergencias se debe siempre buscar áreas de mejora operativa con el objetivo de brindar un mejor servicio a la ciudadanía. Es por eso que constantemente se sugieren y emplean nuevos métodos y criterios en distintas áreas del proceso: desde la recepción de la llamada hasta la resolución del servicio.

El propósito de este análisis fue encontrar mejores criterios para que los operadores tuvieran una mayor facilidad y rapidez en la atención y reacción a emergencias en la Ciudad de México en el menor tiempo posible, sin embargo, es necesario hacer consideraciones en el futuro para el modelo.

Uno de los supuestos cruciales en el modelo que debe ser sometido a revisión a través de los resultados es el conjunto de entrenamiento. El criterio con el que se eligió el ruido de llamadas no relacionadas fue aleatorio con el propósito de minimizar el área bajo la curva (y por tanto las llamadas sugeridas), pero sería posible mejorar este criterio a través de diseños experimentales dentro de la operación del número de emergencias.

Una consideración importante es que el proceso desarrollado, la información utilizada, el modelo y sus resultados son un reflejo de la situación actual del número de emergencias de la Ciudad de México. Si por alguna razón los patrones de conducta de los ciudadanos cambiara respecto a la reacción ante una emergencia, sería una problemática distinta a la que se tendría que enfrentar el sistema. Un primer ejemplo de ello es que actualmente se realiza una campaña para promover el uso del 066, con lo cual se espera aumentar el número de llamadas atendidas, y con ello, el número de veces que distintos ciudadanos reportan el mismo incidente.

Un segundo ejemplo es la incorporación de nuevas tecnologías en el reporte de emergencias. Actualmente es posible descargar una aplicación gratuita que permite reportar emergencias a través de un teléfono celular y cada vez es más frecuente el reporte y atención de emergencias a través de redes sociales. Ello conllevará a distintas problemáticas, en las que sin duda, contar con información precisa, una metodología formal y un modelado estadístico y matemático basado en mediciones específicas permitirá tomar las mejores decisiones para la atención a emergencias y el servicio que se le brinda a los ciudadanos.

Bibliografía recomendada

Para entender más sobre el proyecto Ciudad Segura y la implementación de el número de emergencias 066 dentro de la operación de la atención a emergencias de la Ciudad de México:

- Prieto, R. et al., *Análisis Criminal y uso de las Cámaras de Seguridad en la Ciudad de México*, Buenas prácticas para el análisis delictual en América Latina 2013, p. 110-119.

<http://www.pazciudadana.cl/publicacion/buenas-practicas-para-el-analisis-delictual-en-america-latina-2013/?lang=en>

Para entender más sobre la regresión logística usada en el modelo:

- Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome H.. *The Elements of Statistical Learning*, 2009, p. 119-121.