

Detectando Incidentes de Tránsito en Redes Sociales: un Enfoque Inteligente basado en Twitter vs. Waze

Sebastián Vallejos, Brian Caimmi, Diego Alonso, Álvaro Soria, Luis Berdun

Resumen— La detección y divulgación de incidentes de tránsito se ha visto revolucionada con la aparición de las redes sociales. Esto se debe a que las redes sociales permiten que sus usuarios compartan información de utilidad como son los reportes de incidentes de tránsito. No obstante, en la actualidad existen tanto redes sociales genéricas en donde se comunican diversos temas así como también redes sociales focalizadas en un tópico específico. En este artículo presentamos un estudio comparativo entre un enfoque inteligente que detecta incidentes de tránsito a partir del análisis de publicaciones en la red social genérica Twitter y la red social Waze que se especializa en el tránsito. Llevamos a cabo el estudio comparativo considerando como región de interés a la Ciudad Autónoma de Buenos Aires (CABA), Argentina. Los resultados de este trabajo indican que no hay una supremacía por parte de ninguno de los dos enfoques, por lo que deberían ser considerados enfoques complementarios. Esto se fundamenta en que la proporción de detecciones mutuas, es decir incidentes de tránsito detectados por ambos enfoques, resultó ser considerablemente baja no superando el 12% de los casos. Además, los resultados no evidencian que alguno de los enfoques tienda a anticipar a su similar en la detección de incidentes.

Palabras Clave— Incidentes de Tránsito, Twitter, Waze, Aprendizaje de Máquina, Procesamiento de Lenguaje Natural

I. INTRODUCCIÓN

UNO de los grandes problemas que afrontan las personas en las grandes ciudades es el planeamiento de los viajes urbanos. Estos viajes (por ejemplo, de la casa a la oficina y de vuelta a casa) suelen verse afectados por incidentes de tránsito como accidentes y cortes, convirtiéndolos en una actividad

Sebastián Vallejos, ISISTAN-UNCPBA-CONICET, Facultad de Ciencias Exactas, Campus Universitario, Paraje Arroyo Seco, Tandil, Buenos Aires, Argentina, (e-mail: sebastian.vallejos@isistan.unicen.edu.ar).

Brian Caimmi, ISISTAN-UNCPBA-CONICET, Facultad de Ciencias Exactas, Campus Universitario, Paraje Arroyo Seco, Tandil, Buenos Aires, Argentina, (e-mail: brian.caimmi@isistan.unicen.edu.ar).

Diego Alonso, ISISTAN-UNCPBA-CONICET, Facultad de Ciencias Exactas, Campus Universitario, Paraje Arroyo Seco, Tandil, Buenos Aires, Argentina, (e-mail: diego.alonso@isistan.unicen.edu.ar).

Álvaro Soria, ISISTAN-UNCPBA-CONICET, Facultad de Ciencias Exactas, Campus Universitario, Paraje Arroyo Seco, Tandil, Buenos Aires, Argentina, (e-mail: alvaro.soria@isistan.unicen.edu.ar).

Luis Berdun, ISISTAN-UNCPBA-CONICET, Facultad de Ciencias Exactas, Campus Universitario, Paraje Arroyo Seco, Tandil, Buenos Aires, Argentina, (e-mail: luis.berdun@isistan.unicen.edu.ar).

estresante. Para evitar inconvenientes, las personas comúnmente se informan sobre el estado del tránsito antes de iniciar un viaje. Años atrás, los principales medios para informarse sobre el tráfico eran la televisión y la radio. Sin embargo, recientemente las personas comenzaron a utilizar las redes sociales con este fin. Así, por ejemplo, cuando una persona atestigua un incidente de tránsito suele informarlo a través de las redes sociales.

Tras esta iniciativa, las redes sociales se convirtieron en una importante fuente de información sobre el estado del tráfico para una región determinada. En consecuencia, las redes sociales han inducido nuevos campos de investigación en donde se usa la información de estas redes como fuentes de información complementarias para actividades de planificación urbana [1].

En la literatura existe una variedad de trabajos que se centran en la clasificación de publicaciones de Twitter para la detección de eventos relacionados al tránsito [2], [3], [4], [5]. En general, la clasificación de publicaciones de Twitter es llevada a cabo mediante la utilización de técnicas de Machine Learning. Sin embargo, si bien estos enfoques filtran publicaciones de Twitter relacionadas al tránsito, presentan la limitación de no interpretar automáticamente los incidentes de tránsito detectados.

En consecuencia, otros autores han propuesto diferentes enfoques que tienen por objetivo llevar a cabo la interpretación de incidentes de tránsito reportados vía redes sociales como Twitter¹ o Facebook [6], [7], [8], [9]. No obstante, estos trabajos analizan ciertas cuentas predefinidas que suelen reportar los incidentes de tránsito utilizando estructuras sintácticas simples; facilitando así el proceso de interpretación de los incidentes.

También existen enfoques que primero identifican publicaciones de Twitter que reportan incidentes de tránsito y luego llevan a cabo un proceso de interpretación de incidentes de tránsito a partir de las publicaciones consideradas como relevantes [10], [11]. A pesar de que los procesos que se realizan en estos trabajos permiten obtener información que es de utilidad en el contexto de planificación urbana, la información obtenida no es representada en un mapa. Esta última consideración es de suma importancia para facilitar el planeamiento de viajes diarios en grandes ciudades.

Asimismo, si bien los trabajos de la literatura mencionados analizan publicaciones provistas en redes sociales; hasta donde sabemos, no existe un enfoque que analice publicaciones que se encuentren escritas en Español.

Además de las redes multipropósito como Twitter, surgieron nuevos tipos de redes sociales especializadas en tránsito. En este tipo de redes sociales, los usuarios pueden reportar incidentes de tránsito sobre un mapa, alertando así a otros usuarios. Como caso de ejemplo, se puede mencionar a Waze² que es considerada como la red social más popular dedicada exclusivamente al tránsito.

En este punto, en las redes sociales existen dos fuentes de información distintas sobre el estado del tránsito: las redes sociales genéricas y las redes sociales especializadas en tráfico. Al tener dos fuentes de información surgen algunas incógnitas: ¿En qué fuente circula más información? ¿En cuál se informa antes sobre un incidente? Las redes sociales genéricas poseen más usuarios que las redes sociales especializadas en tránsito. Por esta razón, suelen manejar un volumen de información mucho más grande. Sin embargo, solo una pequeña parte de esta información está relacionada exclusivamente al tránsito. Entonces, ¿cuál de las alternativas es mejor para informarse del tránsito?

En este contexto, el presente trabajo intenta responder estas preguntas mediante un estudio comparativo entre una extensión de nuestro enfoque propuesto en [12] y Waze. El experimento consistió en monitorear incidentes de tránsito de la Ciudad Autónoma de Buenos Aires (Argentina) durante cinco días. Los incidentes detectados mediante el enfoque de [12] se compararon con los reportados en Waze. Los resultados obtenidos concluyen en que el enfoque propuesto y Waze deberían ser considerados complementarios. Por un lado, la proporción de detecciones mutuas, es decir incidentes de tránsito detectados por ambos enfoques, no superó el 12% de los casos. Esto indica que habitualmente los enfoques detectan distintos incidentes de tránsito. Por otro lado, los resultados obtenidos indican que existe un balance en cuanto a las veces y el promedio de tiempo en que los enfoques se anticipan al detectar los mismos incidentes.

El resto del artículo se organiza de la siguiente forma. La Sección II presenta el enfoque que se aplica sobre Twitter. La Sección III introduce los conceptos básicos de Waze. La Sección IV detalla el proceso del estudio comparativo. Por último, la Sección V presenta las conclusiones del trabajo y discute posibles trabajos futuros.

II. ENFOQUE PROPUESTO BASADO EN TWITTER

El primer enfoque a comparar es una extensión de nuestro enfoque propuesto en [12]. El enfoque propuesto detectaba reportes de incidentes de tránsito a partir del análisis de publicaciones de la red social Twitter. Para realizar este trabajo fue necesario extender el enfoque, agregándole la capacidad de geolocalizar los incidentes detectados y ubicarlos sobre un mapa. En la Figura 1 se muestra un esquema conceptual del enfoque. El flujo inicia con la captura de

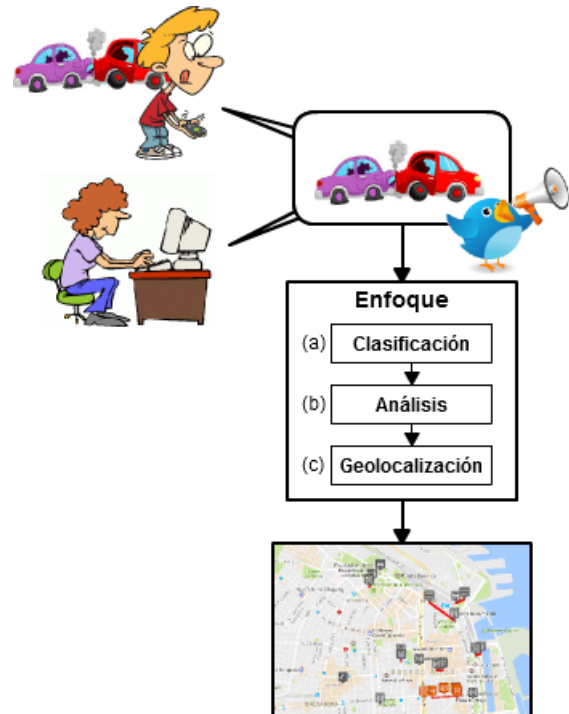


Fig. 1. Enfoque propuesto para la detección de incidentes de tránsito a partir del análisis de publicaciones de Twitter.

publicaciones compartidas en Twitter en tiempo real. En este punto, los usuarios de Twitter generan una gran cantidad de publicaciones por segundo. Por esta razón, el enfoque considera ciertas condiciones a la hora de capturar publicaciones. De esta forma, solo se capturan publicaciones que cumplan estas condiciones. Por ejemplo, publicaciones escritas en determinado idioma, y que contengan determinadas palabras.

Al capturar una publicación, el enfoque la analiza en busca de reportes de incidentes de tránsito. Este análisis consiste en tres etapas. En la primera etapa denominada ‘clasificación’ (a), el enfoque filtra y descarta aquellas publicaciones que no reporten incidentes de tránsito. Para esto, el enfoque primero clasifica cada publicación en ‘relevante’ o ‘irrelevante’ al tópico ‘incidentes de tránsito’ mediante tres clasificadores de texto: *Support Vector Machine* (SVM) [13], [14] empleando dos kernels distintos (Linear Kernel y RBF Kernel) y *Naive Bayes* [15], [16]. Luego, contando con las tres clasificaciones para cada publicación, el enfoque realiza el filtrado. El proceso de filtrado consiste en descartar sólo aquellas publicaciones que hayan sido clasificadas como ‘irrelevante’ por los tres clasificadores. Esta política de filtrado maximiza el *recall*, ya que minimiza las posibilidades de descartar publicaciones que reporten incidentes de tránsito a causa de una mala clasificación.

En la segunda etapa designada ‘análisis’ (b), el enfoque detecta los incidentes y las ubicaciones reportadas en la publicación. Para esto, el enfoque primero reconoce las entidades nombradas en el texto mediante técnicas de *String Matching*. Luego de reconocidas las entidades, se realiza un *análisis sintáctico* mediante reglas sintácticas. Estas reglas relacionan las entidades reconocidas en el texto, sus categorías

² <https://www.waze.com/>

y los conectores que haya entre ellas. Por ejemplo, la regla ‘Calle y Calle → Intersección’ permite reconocer una ubicación. De esta forma, las reglas reconocen ubicaciones, eventos y finalmente incidentes de tránsito. Continuando con el ejemplo, la regla ‘Accidente en Intersección → Incidente’ reconoce un incidente de tránsito a partir de la ubicación reconocida por la regla anterior.

Por último, en la tercera etapa denominada ‘geolocalización’ (c), el enfoque geolocaliza los incidentes detectados y los ubica visualmente sobre un mapa. Geolocalizar un incidente significa obtener las coordenadas geográficas de su ubicación. Por ejemplo, si se detecta un corte en el tramo de una calle comprendido entre otras dos calles, geolocalizar el corte consiste en obtener las coordenadas geográficas de los dos extremos del tramo. Para esto, el enfoque utiliza el servicio de geocodificación provisto por Google [17] que traduce direcciones textuales en coordenadas geográficas. Mediante este servicio, el enfoque traduce la descripción textual de la ubicación indicada en la publicación que reporta al incidente a coordenadas geográficas. Si la geocodificación es exitosa, el enfoque utiliza las coordenadas resultantes para ubicar el incidente sobre un mapa. En cambio, si la geocodificación falla por algún motivo como en el caso de que la ubicación no exista, el enfoque descarta la publicación.

Resumiendo, el enfoque es capaz de analizar las publicaciones compartidas en Twitter en tiempo real y ubicarlos sobre un mapa. Esto permite a los usuarios informarse de manera simple y rápida sobre el estado del tránsito de forma visual. Actualmente, hay un prototipo online que materializa el enfoque. El prototipo (de acceso público en <http://intranet.isistan.unicen.edu.ar/>) se encuentra monitoreando la Ciudad Autónoma de Buenos Aires (CABA). En la página se pueden observar tanto los incidentes detectados a lo largo de la ciudad, como las publicaciones en las que se detectó cada uno de ellos. Con esta información, las personas pueden determinar qué camino tomar para que su movilidad en la ciudad no se vea afectada por los incidentes de tránsito.

III. WAZE

En este trabajo, el enfoque presentado en la Sección II se compara contra Waze. Waze es un sistema de navegación vehicular muy popular en la actualidad. Este sistema brinda a sus usuarios un servicio de planeamiento de rutas en tiempo real. Cuando un usuario necesita viajar desde un punto de la ciudad a otro, Waze busca una ruta que minimice el tiempo de viaje considerando el estado del tránsito en ese momento. Para conocer el estado del tránsito Waze se basa en la colaboración por parte de sus usuarios que son denominados Wazers. Los Wazers pueden colaborar aportando datos o información acerca del estado del tránsito. En la Figura 2 se muestra un esquema conceptual sobre el sistema de colaboración empleado en Waze. Como se observa en la figura, existen dos tipos de colaboración: la colaboración activa (a) y la colaboración pasiva (b).

La colaboración activa (a) ocurre cuando un Wazer reporta un incidente de tránsito. Los usuarios de Waze tienen la

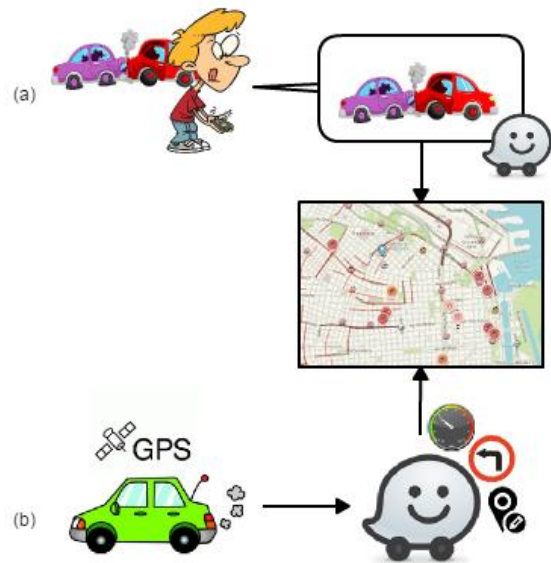


Fig. 2. Sistema de colaboración de la red social Waze.

posibilidad de reportar los incidentes de tránsito que atestiguan. Al reportar un incidente en Waze, se debe indicar el tipo de incidente (accidente, corte, congestión, etc.) y su ubicación geográfica. Nótese que el usuario debe ser testigo del incidente al momento de reportarlo. Por esto, la ubicación del incidente que el usuario indica debe ser cercana a la ubicación del usuario que se determina por el GPS de su dispositivo móvil. Mediante este mecanismo, los usuarios de Waze comparten información acerca del estado del tránsito. Contando con esta información, Waze puede mejorar las estimaciones en los tiempos de viaje y sugerir mejores rutas a sus usuarios. De esta forma, Waze contribuye a formar comunidades de conductores locales que de manera conjunta realizan su aporte para mejorar la calidad de sus viajes diarios.

La colaboración pasiva (b) ocurre cuando un Wazer conduce con la aplicación abierta en su dispositivo móvil. Waze recopila periódicamente datos a partir de los GPS que se encuentran en los dispositivos de sus usuarios como es el caso de los celulares inteligentes. Los datos recopilados proveen información de carácter útil sobre las condiciones de tránsito en diferentes áreas a todos sus usuarios. Por ejemplo, se pueden calcular las velocidades promedio a la que viajan los dispositivos. Esta información resulta útil para estimar duraciones de viajes (al igual que los incidentes reportados en la colaboración activa). Adicionalmente, Waze utiliza los datos recopilados mediante este mecanismo con diversos objetivos: como verificar errores en los mapas; conocer el sentido de las calles; mejorar la estructura vial; o conocer los giros permitidos.

Más allá de la creación de comunidades, Waze posee otras funciones sociales y de geo-gaming que complementan la aplicación. Una de ellas, consiste en un sistema de puntos y logros. Este sistema permite a los usuarios escalar en una jerarquía de rangos en base a la cantidad de incidentes que reportan o al uso pasivo de la aplicación. Escalar en esta jerarquía incentiva a los usuarios a colaborar, aportando más información acerca del estado del tránsito. Otro de los aspectos sociales más interesantes de Waze es su soporte multilingüe y su interacción con redes sociales. Por ejemplo,

un Wazer puede visualizar a sus amigos Wazers de otras redes sociales como Facebook o Twitter en su propio mapa. Estas funciones sociales fomentan la utilización de la aplicación y el crecimiento de la red social interna de Waze.

Resumiendo, el enfoque de Waze consiste en comunidades colaborativas de usuarios que comparten información útil respecto al tránsito. Esta información puede usarse tanto para mantener actualizados los mapas de la ciudad (colaboración pasiva) como para conocer los incidentes de tránsito que ocurren (colaboración activa). Dado que en este trabajo se busca comparar la información acerca de incidentes de tránsito que circula en cada red social, solo la colaboración activa de los Wazers resulta de interés. De esta forma, los incidentes de tránsito reportados por los Wazers se comparan con los incidentes que el enfoque presentado en la Sección II detecta en las publicaciones de Twitter.

IV. ESTUDIO COMPARATIVO

En esta Sección se presenta el proceso llevado a cabo para comparar dos enfoques para la detección de incidentes de tránsito: el enfoque propuesto en este artículo que se basa en la red social Twitter y Waze. Este estudio comparativo focaliza dos criterios de comparación que persiguen diferentes objetivos.

El primer criterio de comparación tiene por objetivo analizar la cobertura de incidentes de tránsito detectados por ambos enfoques. De manera más específica, se desea determinar si el enfoque propuesto que se basa en la red social Twitter es capaz de detectar los mismos incidentes que son reportados en Waze. A su vez, mediante este criterio se puede estipular la proporción de incidentes de tránsito que solamente logran ser interpretados por alguno de los dos enfoques de manera individual. Dicho de otra manera, si el enfoque propuesto detecta incidentes de tránsito que Waze no identifica y viceversa.

El segundo criterio de comparación consistió en identificar experimentalmente cuál de los dos enfoques tiende a detectar incidentes de tránsito con mayor antelación. La antelación es una de las características más importantes de este tipo de enfoques. Si un enfoque divulga incidentes al poco tiempo de que ocurran, el enfoque tiende a ser de mayor utilidad para los usuarios finales que desean evitar cualquier obstáculo durante sus viajes interurbanos. Asimismo, mediante este criterio se puede determinar en cuál de las redes sociales involucradas (Twitter y Waze) se divulgan incidentes de tránsito con mayor anticipación.

El resto de la Sección se estructura de la siguiente forma. La Sub-sección IV-A presenta el conjunto de datos utilizado en el estudio comparativo. La Sub-sección IV-B describe el pre-procesamiento que recibieron estos datos para poder realizar la comparación. La Sub-sección IV-C detalla la descripción y los resultados del proceso comparativo considerando los dos criterios de comparación definidos. Finalmente, la Sub-sección IV-D presenta un apartado que discute los resultados obtenidos.

A. Recolección del Conjunto de Datos

El caso de estudio involucró como región de interés a la Ciudad Autónoma de Buenos Aires (CABA), Argentina. CABA posee la mayor cantidad de habitantes de Argentina y una de las de mayores densidades poblacionales de Sudamérica (14450.8 hab/km² en el año 2010) [18]. Esto da lugar a que CABA usualmente presente un tránsito complicado e incluso caótico. A modo de ejemplo, la ciudad se ubica en la posición 19 del ranking de ciudades con más congestión del mundo [19]. Por lo tanto, asumimos que la frecuencia de incidentes de tránsito que ocurren en esta región es alta, como también lo es la diversidad de usuarios que alertan sobre estos incidentes en las redes sociales Twitter y Waze.

El primer paso de la comparación comprendió la recolección de una muestra de publicaciones de Twitter y alertas de Waze. Para recolectar publicaciones de Twitter se utilizaron dos filtros sobre el flujo público de Twitter. El primer filtro se basa en una lista de palabras clave. Cuando el texto de una publicación presenta alguna de estas palabras la publicación pasa a formar parte del conjunto de datos. En este sentido, utilizamos una lista conformada por 65 palabras clave frecuentemente utilizadas en la jerga referida al tránsito. Palabras como 'tránsito', 'accidente' y 'cortes' son algunos ejemplos utilizados. El segundo filtro aplica una delimitación geoespacial en la que las publicaciones de Twitter que dispongan de las coordenadas geográficas del lugar en que se generó la publicación son consideradas en el conjunto de datos. En este caso, se conformó un cuadrante que delimitara a CABA. Los vértices del cuadrante se corresponden con las siguientes coordenadas: latitud máxima de -34.5329; latitud mínima de -34.7075; longitud máxima de -58.3031; longitud mínima de -58.5324. Se debe tener en cuenta que la API de Twitter utilizada para la recolección sólo provee una proporción del total de publicaciones de Twitter existentes. Específicamente, esta proporción oscila entre el 1% y el 40% de las publicaciones generadas en tiempo real [20], [21]. Con respecto a la recolección de alertas de Waze se empleó un filtro de delimitación geoespacial. En este caso, se utilizó el mismo cuadrante utilizado en la recolección de publicaciones de Twitter. De esta forma, todas las alertas que fueran posicionadas dentro del cuadrante conformaron el conjunto de alertas de Waze.

La recolección de publicaciones de Twitter y de alertas de Waze se realizó durante un lapso de cinco días. El lapso comprendió tanto días laborales como días no laborales en donde controlamos que no existieran reportes programados de manifestaciones o cortes. Al asegurarnos que no existían reportes programados podemos suponer que la región de interés presentaba tránsito de carácter normal. Concretamente, la recolección de datos comenzó un viernes a las 11:00 horas y finalizó un martes a las 11:00 horas. Durante este lapso de tiempo de cinco días, se recolectaron 21075 publicaciones de Twitter y 15983 alertas de Waze.

B. Pre-procesamiento del Conjunto de Datos

Una vez que se tuvieron a disposición los conjuntos de publicaciones de Twitter y de alertas de Waze, se pre-

procesaron estos conjuntos para adecuar los reportes de incidentes al contexto de la comparación. Para realizar este pre-procesamiento, se tuvieron en cuenta una serie de consideraciones. En primer lugar, el enfoque propuesto analizó el conjunto de publicaciones de Twitter (es decir, filtró las publicaciones relevantes al tránsito, las interpretó y las geolocalizó) para detectar los incidentes de tránsito presentes en esta colección. Para realizar el procesamiento, se aclara que se realizaron distintas evaluaciones empíricas con el fin de obtener distintas configuraciones posibles para el enfoque propuesto. Luego, se seleccionó la configuración óptima para realizar el estudio comparativo. En segundo lugar, el análisis comparativo se limitó a dos tipos de incidentes: accidente y cortes. Esto se debe a que ambos enfoques son capaces de detectar estos tipos de incidentes que suelen ser fácilmente localizables y que generalmente abarcan áreas reducidas. Por lo tanto, los incidentes de tránsito que no eran accidentes o cortes (por ejemplo, demoras y manifestaciones) no fueron tenidos en cuenta. En tercer lugar, en la comparación consideramos a cada fecha por separado para poder ver la variación de las detecciones en días laborales (viernes, sábado, lunes y martes) y días no laborales (domingo).

La Tabla I presenta el número de incidentes (accidentes o cortes) detectados por cada uno de los enfoques a lo largo de los días en que tuvo lugar la recolección de datos. Cada uno de los días se detalla mediante su letra inicial. En esta tabla se puede observar que existe una gran disparidad en cuanto al número de incidentes que detectó cada enfoque. Por un lado, el enfoque propuesto basado en Twitter detectó 303 incidentes mientras que Waze reconoció 706 incidentes de tránsito. Estos primeros valores sugieren que Waze identifica más del doble de incidentes de tránsito que el enfoque propuesto. No obstante, al realizar un análisis detallado, no todos los incidentes se encontraban en las condiciones necesarias para realizar la comparación. Como primera consideración, algunos incidentes detectados por el enfoque propuesto fueron geolocalizados por fuera del cuadrante delimitador de CABA. Como segunda consideración, era necesario que los incidentes tuvieran cierto grado de confianza. En efecto, un incidente presenta un grado aceptable de confianza si la fuente que lo reporta es considerada fiable por otros usuarios, o si varios usuarios reportan el mismo incidente. Sin embargo, no todos los incidentes involucrados cumplían con este requisito.

Para continuar con el pre-procesamiento, se adecuaron los incidentes a la comparación mediante la utilización de dos criterios de aceptación. El primer criterio de aceptación

establece que todos los incidentes detectados se encuentren localizados dentro de una misma región espacial. En consecuencia, todos los incidentes de tránsito detectados por el enfoque propuesto que fueran localizados fuera del cuadrante delimitador no se tuvieron en cuenta durante el resto del análisis comparativo. El segundo criterio de aceptación determina que los incidentes detectados deben presentar cierto grado de confianza para ser tenidos en cuenta. Por ejemplo, el hecho de que un incidente sea reportado por una única publicación de Twitter no tiene el mismo grado de confianza que el de un incidente reportado múltiples veces por distintos usuarios. Por lo tanto, los incidentes de tránsito detectados por el enfoque propuesto que estuvieran asociados a distintos reportes se mantuvieron para realizar la comparación, mientras que los que tuvieran asociada una única publicación fueron descartados. Algo similar se llevó a cabo con las alertas de Waze. Para este caso, los metadatos de las alertas de Waze ofrecen ciertos atributos que facilitan esta tarea. Uno de estos atributos se denomina '*Confidence*'. Este atributo determina en un rango [0; 10] el valor de confianza de la alerta de tránsito basándose en la reputación del usuario que la creó. Asimismo, otro atributo denominado '*nThumbsUp*' determina el número de validaciones positivas proporcionadas por otros usuarios que circularon cerca del incidente. Teniendo en cuenta estos dos atributos, cualquier alerta que tuviera al menos un valor mayor a cero para el atributo '*Confidence*' o '*nThumbsUp*' fue considerada para la comparación.

La Tabla II muestra el número de incidentes de tránsito detectados por cada enfoque al considerar a aquellos incidentes que hayan satisfecho los criterios de aceptación. A partir de esta tabla se puede ver que la cantidad de incidentes detectados por el enfoque propuesto pasó de un total inicial de 303 incidentes a 244 incidentes lo que indica una reducción del 19.47%). Además, la cantidad de incidentes detectados por Waze se redujo notablemente de 706 incidentes iniciales a 463 incidentes lo que expresa una reducción del 34.41%.

El siguiente paso del pre-procesamiento comprendió la eliminación de casos en donde existan repeticiones de incidentes. Una repetición de incidentes ocurre cuando existen al menos dos incidentes del mismo tipo cuyas ubicaciones tengan una distancia cercana, por lo que se los puede interpretar como el mismo incidente. Por ejemplo, en Waze es común que distintos usuarios reporten un mismo incidente con distintas ubicaciones cercanas, generando así redundancia. Esta redundancia desvirtúa los resultados del estudio

TABLA I
INCIDENTES DETECTADOS POR EL ENFOQUE PROPUESTO Y POR LA RED SOCIAL WAZE

Enfoque	Incidente	Día					Total
		V	S	D	L	M	
Enfoque Propuesto	Accidentes	21	22	15	42	13	113
	Cortes	42	27	24	55	42	190
	Total	63	49	39	97	55	303
Waze	Accidentes	139	97	50	95	35	416
	Cortes	74	54	19	79	32	258
	Total	213	151	69	174	67	706

TABLA II
INCIDENTES QUE SATISFACEN LOS CRITERIOS DE ACEPTACIÓN PARA REALIZAR LA COMPARACIÓN

Enfoque	Incidente	Día					Total
		V	S	D	L	M	
Enfoque Propuesto	Accidentes	15	20	12	28	5	80
	Cortes	36	23	20	46	39	164
	Total	51	43	32	74	44	244
Waze	Accidentes	77	53	19	42	14	205
	Cortes	74	54	19	79	32	258
	Total	151	107	38	121	46	463

comparativo, por lo que debe ser eliminada. En consecuencia, en cada caso de repetición de incidentes, el incidente que presente una mayor antelación en su tiempo de reporte es mantenido y los incidentes redundantes son eliminados.

Para llevar a cabo el procedimiento de eliminación de incidentes repetidos, se analizaron los incidentes utilizando diferentes radios de distancia. Al usar varios radios se pueden agrupar los incidentes por cercanía. Para este caso, se variaron los radios de distancia entre las ubicaciones de los incidentes en 50 mt, 100 mt, 200 mt y 300 mt. Los radios abarcaron distancias relativamente cercanas (50 mt y 100 mt) inferiores al largo en metros de una cuadra en CABA para cuando se conoce con exactitud la ubicación de los incidentes. En cambio, se emplearon distancias más largas (200 mt y 300 mt) para soportar los casos donde existan reportes de un mismo incidente del que se conoce su zona y no su ubicación precisa.

En la Tabla III se puede apreciar el estado del conjunto de datos luego de aplicar la eliminación de incidentes repetidos. Se debe aclarar que el estado del conjunto de datos antes de la eliminación de repeticiones se detalla en negrita junto con un guion (‘-’) para la columna radio. A partir de esta tabla, se puede determinar que existen diferencias entre la cantidad de incidentes repetidos que procesan ambos enfoques. A medida que se aumenta el valor del radio se disminuye las detecciones de ambos enfoques ya que se van eliminando repeticiones. Sin embargo, la disminución de repeticiones es más notoria en el caso de Waze que en el caso del enfoque propuesto. Pongamos por caso a los cortes detectados en el día viernes. Con un radio de 50 metros, la cantidad de cortes detectados por el enfoque propuesto se reduce de 36 a 34. Mientras que en el caso de Waze se reduce de 74 a 33.

TABLA III
DISTRIBUCIONES DE INCIDENTES AL REALIZAR LA ELIMINACIÓN DE INCIDENTES REPETIDOS

Enfoque	Incidente	Radio [mt]	Día				
			V	S	D	L	M
Enfoque Propuesto	Accidentes	-	15	20	12	28	5
		50	14	12	10	25	5
		100	13	12	10	23	5
		200	12	12	10	22	5
		300	12	11	10	21	5
	Cortes	-	36	23	20	46	39
		50	34	23	20	42	33
		100	32	22	20	42	33
		200	28	15	18	30	24
		300	28	14	15	28	20
Waze	Accidentes	-	77	53	19	42	14
		50	60	44	17	37	12
		100	50	38	16	34	12
		200	47	34	15	32	11
		300	41	32	15	32	11
	Cortes	-	74	54	19	79	32
		50	33	33	16	45	19
		100	33	32	14	42	18
		200	29	26	8	33	16
		300	28	22	8	30	15

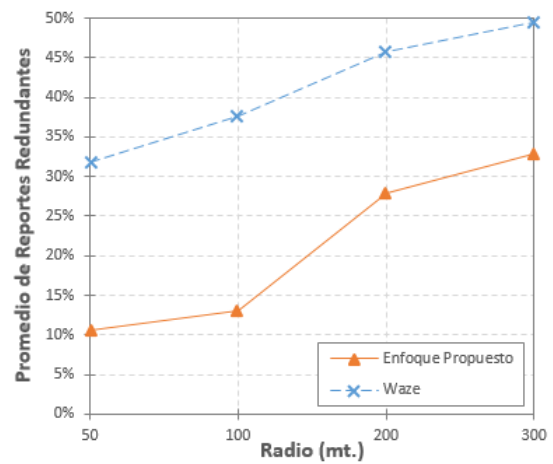


Fig. 3. Promedio de reportes redundantes para el enfoque propuesto y para Waze al considerar radios de 50 mt, 100 mt, 200 mt y 300 mt.

La Figura 3 describe el promedio de reportes redundantes que tiene el conjunto de datos utilizados para cada uno de los enfoques. En el eje horizontal se observan los distintos radios considerados para el experimento (50 mt, 100 mt, 200 mt y 300 mt) mientras que en el eje vertical se detalla el promedio de reportes redundantes para cada radio utilizado. En la Figura 3 se puede visualizar el hecho de que en todos los casos de radios el enfoque propuesto posee una menor cantidad de incidentes repetidos. Cuando se considera un radio de 50 mt y de 100 mt, la diferencia llega al 21.09% y al 24.47%, respectivamente. Luego, para el caso de 200 mt y 300 mt, la diferencia comienza a reducirse gradualmente llegando a 17.92% y 16.67%. A partir de esos datos, podemos decir que el radio de 100 metros puede ser considerado como un punto de inflexión debido a que en radios más altos (200 y 300 metros) se tiende a agrupar una gran proporción de incidentes distintos como repeticiones. Además, dado que en CABA las cuadras tienen un largo de 100 metros, si se buscaran detecciones mutuas en radios mayores a este valor es posible que se agrupen distintos incidentes de tránsito. En resumen, se puede afirmar que Waze presenta en todos los casos una mayor proporción de incidentes repetidos y que el radio de 100 metros es acorde para realizar el estudio comparativo.

El hecho de que Waze presente una mayor proporción de incidentes repetidos se debe a ciertos aspectos como el hecho de que ante un incidente de tránsito cada usuario de la red puede generar una nueva alerta o brindarle confianza a una alerta del incidente ya existente mediante una validación positiva. Sin embargo, analizando visualmente los incidentes en un mapa se concluyó que existe una tendencia por parte de los usuarios de Waze a generar reportes nuevos para un mismo incidente. La Figura 4 presenta en un mapa que para un mismo incidente reportado en el día viernes del conjunto existen una gran cantidad de alertas de Waze (Sub-figura b) con tiempo de reportes similares y ubicaciones cercanas, mientras que para ese incidente el enfoque propuesto localizó el incidente en un único lugar (Sub-figura a). Esto se debe a que el proceso de geolocalización del enfoque propuesto siempre asigna las mismas coordenadas geográficas para distintos reportes que expresen la misma ubicación aunque en sus textos la ubicación

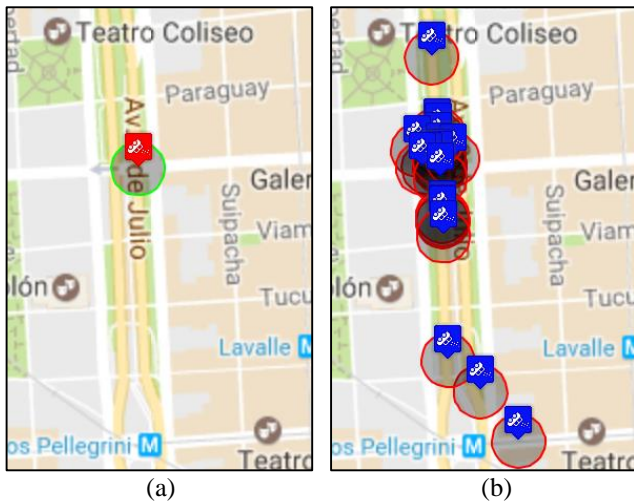


Fig. 4. Un caso de ejemplo del día viernes del conjunto de datos en donde se repiten incidentes. La Sub-figura (a) muestra el incidente detectado por el enfoque propuesto. La Sub-figura (b) ilustra las alertas de Waze generadas para un mismo incidente.

sea escrita de diferentes formas. De manera contraria, los usuarios de Waze generaron muchas alertas con diferente tiempo de reporte y ubicación para el mismo incidente. Se aclara que las ubicaciones son diferentes debido a que Waze utiliza la posición GPS de los usuarios para indicar la posición de los incidentes reportados.

C. Comparación de Incidentes Detectados

Una vez que se finalizó con el pre-procesamiento del conjunto de datos, se continuó con el primer criterio de comparación que se centra en analizar la cobertura de incidentes detectados por ambos enfoques. Por lo tanto, se llevó a cabo un análisis sobre el conjunto de incidentes sin repeticiones considerando separadamente los tipos de incidentes involucrados (es decir, accidentes y cortes) y un radio de 100 metros. Específicamente, el conjunto de incidentes utilizado se desprende de las filas de la Tabla III en donde el radio fuera de 100 metros.

Con el objetivo de analizar el conjunto de reportes conseguidos y determinar la distribución de detecciones, se efectuaron los siguientes pasos. Primero, por cada tipo de incidente, se presentaron los incidentes detectados por cada enfoque de manera superpuesta en un mismo mapa de CABA. Luego, por cada incidente reconocido por el enfoque

propuesto se calculó la distancia en metros hacia todos los incidentes reconocidos por Waze. Cuando la distancia hacia algún incidente reconocido por Waze fuera igual o menor al radio de 100 metros establecido, los incidentes fueron considerados como una detección mutua. Dicho de otra manera, en estos casos ambos enfoques detectaron el mismo incidente de tránsito. Por el contrario, los incidentes que no participaron de alguna detección mutua fueron considerados como detecciones individuales. Al finalizar, se hizo un recuento para determinar el número de detecciones propias del enfoque propuesto, el número de detecciones propias de Waze y el número de detecciones mutuas entre ambos enfoques.

La distribución resultante de incidentes detectados por el enfoque propuesto y Waze considerando un radio de 100 metros se detalla en la Tabla IV. La primera columna determina el tipo de incidente, es decir accidente o corte. La segunda columna especifica el tipo de detección que puede ser tanto individual (denotada con el nombre del enfoque correspondiente) o conjunta para el caso de las detecciones mutuas. La tercera columna presenta la cantidad de incidentes involucrados en cada uno de los días tenidos en cuenta en el experimento. La cuarta columna indica el total de incidentes detectados para cada caso considerando todas las fechas. La quinta y última columna presenta la proporción de detecciones totales para cada caso.

Analizando las Tabla IV se puede apreciar que la cantidad de detecciones mutuas fue mucho más baja que la cantidad de incidentes detectados individualmente por cualquiera de los enfoques. Esto ocurrió con proporciones similares tanto para el caso de los accidentes (11.74%) como para el caso de los cortes (9.72%). A nuestro parecer, que existan pequeñas proporciones de detecciones mutuas induce a que los reportes de tránsito que circulan en Twitter y en Waze no suelen cubrir los mismos incidentes.

Centrándonos en los resultados para los accidentes de la Tabla IV, Waze identificó una cantidad de incidentes considerablemente mayor que el enfoque propuesto. Específicamente, Waze logró una proporción de detecciones (64.32%) que duplica a la proporción conseguida por el enfoque propuesto (23.94%). Sin embargo, gran parte de esta diferencia se debe a la propiedad de ubicación de accidentes en Waze que fue descrita anteriormente en la Figura 4. Si bien

TABLA IV
DISTRIBUCIÓN DE INCIDENTES DETECTADOS INDIVIDUALMENTE Y MUTUAMENTE POR EL ENFOQUE PROPUESTO Y WAZE AL CONSIDERAR UN RADIO DE 100 METROS

Tipo de Incidente	Tipo de Detección	Día					Detecciones Totales	Proporción de Totales
		V	S	D	L	M		
Accidentes	Enfoque Propuesto	11	8	8	21	3	51	23.94 %
	Waze	48	33	14	32	10	137	64.32 %
	Mutuas	4	9	4	4	4	25	11.74 %
	Total	63	50	26	57	17	213	100.00 %
Cortes	Enfoque Propuesto	26	22	17	39	31	135	46.88 %
	Waze	27	32	11	39	16	125	43.40 %
	Mutuas	12	0	6	6	4	28	9.72 %
	Total	46	34	34	84	51	288	100.00 %

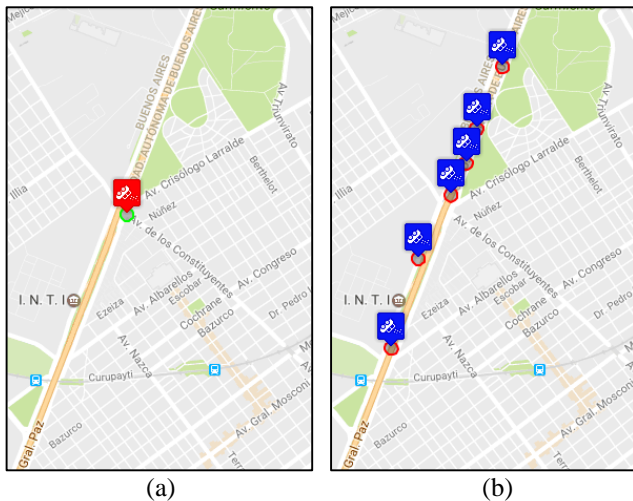


Fig. 5. Repetición de incidentes en el caso de una autopista de CABA. La Sub-figura (a) detalla el accidente detectado por el enfoque propuesto mientras que la Sub-figura (b) muestra los accidentes detectados por Waze.

se llevó a cabo la eliminación de repeticiones, en algunos casos el proceso automático utilizado es incapaz de detectar ciertos incidentes repetidos. Pongamos por ejemplo el fenómeno ilustrado en la Figura 5. Es este caso, ante un embotellamiento por un accidente en una autopista, varios usuarios en diferentes lugares del embotellamiento pueden reportar la situación a grandes distancias. De esta forma, al no poder eliminar estas repeticiones de incidentes, se incrementa la cantidad de accidentes que identifica Waze.

Sin embargo, para el caso de los resultados sobre los cortes de la Tabla IV, la cantidad de incidentes que detectaron ambos enfoques de manera individual resultaron ser similares. En particular, el enfoque propuesto logró una proporción de detecciones (46.88%) que superó levemente a la proporción alcanzada por Waze (43.40%). La justificación de que ocurra esto se basa en el hecho de que al especificar un corte en Waze resulta indispensable detallar las ubicaciones de inicio y de fin. Además, si la información suministrada de los cortes es correcta, los demás usuarios de la red tienden a validar estos reportes y no generar nuevos.

La cantidad de incidentes detectados es mayor en los días laborales que en los no laborales. Al sumar las cantidades de detecciones totales por cada día se obtienen los siguientes valores: el viernes se detectaron 128 incidentes, el sábado 104, el domingo 60, el lunes 141 y el martes 68. Se debe resaltar que el día domingo que es considerado como día no laboral por definición general hubo la menor cantidad de detecciones. Esto puede deberse a que en estos días suele aliviarse el tránsito en CABA. Asimismo, a partir de estos valores se puede calcular las reducciones proporcionales de detecciones entre los demás días (el sábado incluido por ser día de trabajo de media jornada) y el domingo. Por ejemplo, contrastando el lunes (141 incidentes) y el domingo (60 incidentes) se obtiene la reducción máxima de incidentes que fue de 57.44%. La comparación de los demás días contra el domingo devuelve reducciones de detecciones que superan el 42.30% a excepción del día martes. Al comparar el martes y el domingo se obtiene el valor de reducción de incidentes más bajo

(11.76%) debido a que el martes sólo se detectaron incidentes hasta las 11:00 am. Esto concluye en que los días laborales suelen tener una mayor cantidad de detecciones de incidentes de tránsito que los días no laborales.

D. Comparación de los Tiempos de Reporte de los Incidentes que Ambos Enfoques Detectaron Mutuamente

A partir del conjunto de incidentes de tránsito detectados por ambos enfoques cuya distancia sea cercana (es decir, lo que denominamos detecciones mutuas), el siguiente paso en la comparación tiene por objetivo determinar qué enfoque identificó dichos incidentes con mayor antelación. Con tal motivo, la comparación se basó en contrastar los tiempos de reportes de los incidentes que forman el conjunto de detecciones mutuas. Para llevar a cabo esto, para cada día se agruparon los incidentes involucrados en una detección mutua considerando un radio de 100 metros. Por cada caso de detección mutua, se determinó el enfoque que haya detectado al incidente involucrado con mayor anticipación y el lapso de tiempo existente hasta que el enfoque contrario detectara el mismo incidente. Cabe aclarar que estos lapsos de tiempo permiten determinar la diferencia de tiempo existente para que ambos enfoques se percaten de un mismo incidente. Tomando en cuenta estos lapsos de tiempo, se calculó el promedio de antelación total en horas, minutos y segundos, con el que cada enfoque anticipa al enfoque contrincante.

La Tabla V detalla los resultados referidos a las antelaciones entre ambos enfoques. En la primera columna se especifica el tipo de incidente considerado. En la segunda columna se detalla el día involucrado. En la tercera columna se muestra la cantidad de antelaciones (es decir, la cantidad de veces que un enfoque identificó un incidente antes que su oponente). En la cuarta columna se presenta el promedio en horas, minutos y segundos, con el que un enfoque identifica un incidente antes que el otro enfoque. Se debe notar que una celda con guion significa que el enfoque involucrado no logró enterarse de ningún incidente antes que el otro enfoque para una fecha específica, por lo que no tiene asociado un promedio de antelación.

Considerando el caso de los accidentes, Waze tuvo una mayor cantidad de antelaciones en tres de cinco días, mientras

TABLA V
CANTIDAD DE ANTELACIONES Y PROMEDIO DE ANTELACIONES DE AMBOS ENFOQUES UTILIZANDO UN RADIO DE 100 METROS

Tipo de Incidente	Día	Antelaciones		Promedio de Antelación	
		Enfoque Propuesto	Waze	Enfoque Propuesto	Waze
Accidentes	V	1	1	0:13:41	0:22:19
	S	1	3	0:00:15	0:46:15
	D	0	2	-	0:29:17
	L	2	0	3:16:58	-
	M	0	2	-	0:57:53
	Total	4	8	1:41:58	0:41:55
Cortes	V	1	5	5:12:04	1:33:13
	S	-	-	-	-
	D	3	0	2:16:54	-
	L	1	2	1:12:29	0:17:23
	M	1	1	0:43:05	0:52:51
	Total	6	8	2:19:44	1:09:12

que el enfoque propuesto superó a Waze en cantidad de antelaciones en una sola ocasión. Sin embargo, el promedio de antelación total sugiere que el día lunes el enfoque propuesto tuvo un grado de antelación significativamente mayor que el grado presentado por Waze, alcanzando 3 horas de antelación. Al analizar las detecciones mutuas involucradas en estos casos, se descubrió que pertenecían a reportes comprendidos entre las 4 am y 8 am. Según nuestro punto de vista, pudimos determinar que en el rango de horas que comprende el inicio de las jornadas comienzan a circular reportes de tránsito por Twitter mientras que Waze identifica los mismos reportes luego de lapso de tiempo considerable.

Respecto al caso de los cortes, ambos enfoques lograron anticiparse en una cantidad de veces similar. Concretamente, el enfoque propuesto tuvo 6 anticipaciones y Waze 8. Además, analizando los promedios de antelación se ve que el enfoque propuesto logró promedio mayores que los conseguidos por Waze en tres días (viernes, domingo y lunes). Asimismo, el enfoque propuesto anticipó a Waze en todas las detecciones mutuas del día domingo. Se debe recalcar que el enfoque logró antelaciones superiores a 5 horas en el día viernes. En esta ocasión, el corte involucrado fue reportado vía Twitter alrededor de las 11:00 am, mientras que Waze identificó el mismo corte un poco antes de las 16:30 pm.

Finalmente, si se analizan los promedios de antelación para los 5 días, se debe enfatizar el hecho de que el enfoque propuesto tiene un promedio de antelación considerablemente mayor a Waze. Esto ocurrió tanto para accidentes como para cortes. En el primer caso, la diferencia de los promedios fue de 1 hora. En el segundo caso, la misma diferencia se extendió hasta 1 hora y 10 minutos.

E. Discusiones

A partir de la experimentación realizada, hemos comparado la cobertura de detecciones de incidentes de tránsito del enfoque inteligente propuesto que se basa en la red social Twitter y la misma cobertura lograda por Waze. No obstante, decidir qué enfoque alcanza una mayor cobertura no resulta ser una tarea trivial. Por lo tanto, consideramos dos criterios de comparación.

El primer criterio considera la cantidad de incidentes que los enfoques detectan tanto individualmente (incidentes solamente detectados por uno de los enfoques) como conjuntamente (incidentes detectados mutuamente por ambos enfoques). A partir de los resultados presentados en la Subsección IV-C se puede concluir que los enfoques deberían ser considerados complementarios en cuanto a la detección de incidentes. Si bien Waze detectó el doble de accidentes que el enfoque propuesto, muchas de estas detecciones eran repeticiones de incidentes que no pudieron ser eliminados utilizando el mecanismo basado en radios de distancia. A su vez, por lo percibido durante la experimentación, en muy pocos casos los enfoques identificaron mutuamente los mismos incidentes. Con esta evidencia se puede afirmar que las redes sociales Twitter y Waze no suelen divulgar los mismos incidentes de tránsito.

Por otro lado, el segundo criterio toma en cuenta el tiempo

de antelación con el que los enfoques detectan los mismos incidentes. Los resultados ofrecidos en la Sub-sección IV-D determinaron que ninguno de los enfoques logró anticiparse siempre a su oponente. Waze tuvo una mayor cantidad de anticipaciones frente al enfoque propuesto. Este hecho fue notorio en el caso de los accidentes. Sin embargo, en el caso de los cortes la diferencia fue solamente de dos antelaciones. A pesar de este hecho, el promedio de antelación logrado por el enfoque propuesto fue mucho mayor que el promedio de antelación conseguido por Waze para los dos tipos de incidentes considerados. A nuestro parecer, a partir de estas argumentaciones no es posible determinar acertadamente qué enfoque posee mayor anticipación al detectar incidentes de tránsito.

Otro punto interesante a discutir es que pudimos corroborar que los días laborales presentan una mayor proporción de detecciones de incidentes de tránsito que los días no laborales. Esto puede deberse tanto a que en los días no laborales no suele haber el mismo flujo de tránsito que en un día laboral como así a al hecho de que también se reduce la cantidad de Wazers activos. Hay que recordar que los Wazers deben estar 'in situ' para poder reportar el incidente de tránsito.

En conclusión, el estudio comparativo llevado a cabo nos muestra que ninguno de los dos enfoques aventaja ampliamente a su contrincante. A nuestro juicio, los enfoques deberían ser considerados complementarios. La justificación a esto se basa en que ambos enfoques detectan una mayor cantidad de incidentes de forma individual que de forma conjunta. Esto es evidente en la Tabla IV en donde las proporciones de detecciones mutuas no superaron el 12% de las detecciones. En cuanto a tiempo de antelación, Waze logró una mayor cantidad de anticipaciones pero el enfoque propuesto cuando anticipa a Waze lo hace con lapsos de tiempo considerablemente mayores.

V. CONCLUSIÓN

En este trabajo presentamos un estudio comparativo entre dos enfoques orientados a la detección y divulgación de incidentes de tránsito. El primer enfoque analiza publicaciones de la red social Twitter mediante técnicas de *Machine Learning* y *Procesamiento de Lenguaje Natural* para identificar incidentes de tránsito. Sin embargo, al ser una red social genérica, las publicaciones compartidas involucran cualquier tema de interés y sólo una pequeña porción se corresponde al tránsito. El segundo enfoque es la red social Waze que se especializa en la divulgación del estado del tránsito. A diferencia de Twitter, Waze solamente contiene información referida al tránsito. El estudio comparativo que involucró a estos dos enfoques se basó en analizar por un lado la cobertura en cuanto a incidentes de tránsito detectados y por otro lado el grado de antelación de las detecciones.

A partir de este estudio comparativo, se obtuvieron resultados alentadores que indican que los enfoques deberían ser considerados complementarios. Esto se debe a que los enfoques no suelen identificar los mismos incidentes a la vez, sino que cada uno habitualmente detecta incidentes que el otro enfoque no. Además, divisamos que ninguno de los enfoques

tiene un mayor grado de anticipación que el otro. Finalmente, percibimos que luego de eliminar incidentes repetidos (es decir, reportes de igual tipo con ubicación y tiempo similares), Waze sigue presentando una gran cantidad de incidentes que pueden ser considerados redundantes. Asimismo, es necesario notar que la API de Twitter utilizada no garantiza la entrega del 100% de las publicaciones en tiempo real. En consecuencia, es posible que se omitan publicaciones que podrían informar incidentes de tránsito. A pesar de esta limitación, el enfoque propuesto nos brinda resultados competitivos con los arrojados por Waze, lo que resalta la potencialidad de nuestro enfoque.

Con la evidencia empírica de que ambos enfoques deberían ser considerados complementarios, se planean continuar con diferentes líneas de investigación en este campo. En primer lugar, a partir de la evaluación preliminar se pudo concluir que los enfoques no suelen identificar los mismos incidentes, por lo que resultaría interesante poder combinar la información que brinda cada enfoque de manera separada. Al desarrollar un nuevo enfoque que combine los reportes de incidentes de varias fuentes de información se podría mejorar la cobertura de detecciones de incidentes. De esta forma, los usuarios finales percibirían incidentes de tránsito que no se les revelarían si solamente dependieran de uno de estos enfoques. En segundo lugar, si este tipo de análisis se ejecuta continuamente en tiempo real, podría servir como método de validación de incidentes. En otras palabras, esto permitiría verificar que realmente los incidentes de tránsito hayan ocurrido. En tercer lugar, suponer un radio de 100 metros para realizar la comparación de detecciones puede resultar controversial. Una posible alternativa para mejorar el experimento sería establecer el radio de manera variable dependiendo de la velocidad máxima permitida de la vía en donde se reporta un incidente. En cuarto y último lugar, nos interesa repetir el experimento con publicaciones de Twitter y alertas de Waze que abarquen un lapso de tiempo mayor. Además, con una mayor cantidad de datos a evaluar, se podría analizar varios aspectos como los rangos horarios en donde los reportes suelen generarse con mayor frecuencia, las zonas de CABA en donde es habitual que ocurran incidentes de tránsito, e incluso las variaciones que sufre el curso normal del tránsito cuando ocurren eventos especiales como paros o movilizaciones multitudinarias.

REFERENCIAS

- [1] Frias-Martinez, V. & Frias-Martinez, E., "Spectral clustering for sensing urban land use using Twitter activity", *Engineering Applications of Artificial Intelligence*, Elsevier, 2014, 35, 237-245.
- [2] Carvalho, S. F. L. d. & others, "Real-time sensing of traffic information in twitter messages", *Proceedings of the 4th Workshop on Artificial Transportation Systems and Simulation ATSS at IEEE ITSC 2010*, 2010.
- [3] Schulz, A.; Ristoski, P. & Paulheim, H., "I see a car crash: Real-time detection of small scale incidents in microblogs", *Extended Semantic Web Conference*, 2013, 22-33.
- [4] D'Andrea, E.; Ducange, P.; Lazzarini, B. & Marcelloni, F., "Real-time detection of traffic from twitter stream analysis", *IEEE Transactions on Intelligent Transportation Systems*, IEEE, 2015, 16, 2269-2283.
- [5] Kuflik, T.; Minkov, E.; Nocera, S.; Grant-Muller, S.; Gal-Tzur, A. & Shoor, I., "Automating a framework to extract and analyse transport related social media content: The potential and the challenges", *Transportation Research Part C: Emerging Technologies*, Elsevier, 2017, 77, 275-291.
- [6] Endarnoto, S. K.; Pradipta, S.; Nugroho, A. S. & Purnama, J., "Traffic condition information extraction & visualization from social media twitter for android mobile application", *Electrical Engineering and Informatics (ICEEI)*, 2011 International Conference on, 2011, 1-4.
- [7] Kosala, R.; Adi, E. & others, "Harvesting real time traffic information from Twitter", *Procedia Engineering*, Elsevier, 2012, 50, 1-11.
- [8] Anantharam, P.; Barnaghi, P.; Thirunarayan, K. & Sheth, A., "Extracting city traffic events from social streams", *ACM Transactions on Intelligent Systems and Technology (TIST)*, ACM, 2015, 6, 43.
- [9] Albuquerque, F. C.; Casanova, M. A.; Lopes, H.; Redlich, L. R.; de Macedo, J. A. F.; Lemos, M.; de Carvalho, M. T. M. & Renso, C., "A methodology for traffic-related Twitter messages interpretation", *Computers in Industry*, Elsevier, 2016, 78, 57-69.
- [10] Wanichayapong, N.; Pruthipunyaskul, W.; Pattara-Atikom, W. & Chaovalit, P., "Social-based traffic information extraction and classification", *ITS Telecommunications (ITST)*, 2011 11th Int. Conf., 2011, 107-112.
- [11] Gu, Y.; Qian, Z. S. & Chen, F., "From Twitter to detector: Real-time traffic incident detection using social media data", *Transportation Research Part C: Emerging Technologies*, Elsevier, 2016, 67, 321-342.
- [12] Caimmi, B.; Vallejos, S.; Berdun, L.; Soria, Á.; Amandi, A. & Campo, M., "Detección de incidentes de tránsito en Twitter", *Biennial Congress of Argentina (ARGENCON)*, 2016 IEEE, 2016, 1-6.
- [13] Cortes, C. & Vapnik, V., "Support-vector networks", *Machine learning*, Springer, 1995, 20, 273-297.
- [14] Vapnik, V., "The nature of statistical learning theory", *Springer Science & Business Media*, 2013.
- [15] John, G. H. & Langley, P., "Estimating continuous distributions in Bayesian classifiers", *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, 1995, 338-345.
- [16] Schütze, H., "Introduction to Information Retrieval", *Proceedings of the international communication of association for computing machinery conference*, 2008.
- [17] "The Google Maps Geocoding API", <https://developers.google.com/maps/documentation/geocoding/intro>, 2016, [Accedido: 7 de Abril 2017]
- [18] "Buenos Aires", [Es.wikipedia.org](https://es.wikipedia.org/wiki/Buenos_Aires), 2017, [Online], Disponible: ["https://es.wikipedia.org/wiki/Buenos_Aires"](https://es.wikipedia.org/wiki/Buenos_Aires), [Accedido: 7 de Abril de 2017].
- [19] "TomTom Traffic Index", 2017, [Online], Disponible: ["https://www.tomtom.com/en_gb/trafficindex/city/buenos-aires"](https://www.tomtom.com/en_gb/trafficindex/city/buenos-aires), [Accedido: 7 de Abril 2017].
- [20] "Twitter Firehose vs. Twitter API: What's the difference and why should you care? - BrightPlanet", BrightPlanet, 2017, [Online], Disponible: <https://brightplanet.com/2013/06/twitter-firehose-vs-twitter-api-whats-the-difference-and-why-should-you-care/>, [Accedido: 10 de Abril de 2017].
- [21] Bifet, A., Holmes, G., Pfahringer, B., & Gavaldá, R., "Detecting Sentiment Change in Twitter Streaming Data". In *WAPA* (pp. 5-11), October 2011.