

FILTRADO DE DATOS TWITTER PARA CAPTURAR PATRONES DE MOVILIDAD DE UN ÁREA METROPOLITANA. APLICACIÓN AL ÁREA DE VALENCIA

Carmen Zornoza-Gallego¹, Julia Salom-Carrasco¹, Ignacio Maestro-Cano¹

¹ Instituto de Desarrollo Local, Departamento de Geografía, Universidad de Valencia. Avda. Blasco Ibáñez, 28. 46010. Valencia, España. carmen.zornoza@uv.es; Julia.Salom@uv.es; igmaeca@gmail.com

RESUMEN

Este trabajo pretende aproximarnos al empleo de datos provenientes de Big Data para obtener información sobre la movilidad de los habitantes de un área determinada. En este caso se empleará la red social Twitter por la amplia disponibilidad de datos georreferenciados.

La principal aportación se basa en el estudio de las características de la fuente y en su validez para estudiar la movilidad de un área. Se elabora una propuesta metodológica para el filtrado de los datos con el fin de adecuarlos a la obtención patrones de movilidad empleando Sistemas de Información Geográfica (SIG).

La aplicación práctica se centra en la provincia de Valencia como una primera aproximación al estudio del Área Metropolitana de Valencia. La validación de la fuente de datos se ha centrado en la obtención del lugar de residencia de los usuarios y en su comparación con el padrón municipal. Los resultados obtenidos presentan una fuerte correlación, indicando que la metodología es adecuada y que las posibilidades de uso de esta información son amplias.

Palabras clave: Big Data; Movilidad; Sistemas de Información Geográfica.

ABSTRACT

The aim of this paper is to study the possible use of data from Big Data to obtain information about the mobility of a community. The social network we have chosen is Twitter due to the great range and easy accessible geo-referenced data it provides.

The main contribution of the investigation is based on the analysis of the source characteristics and its possibilities for being used as valid information for a mobility study. A methodological proposal was created to filter raw data with the objective to obtain any patterns of mobility, utilizing Geographical Information Systems.

In the initial approach we collected data of the entire province of Valencia in order to allow a further study of the Valencian Metropolitan Area. We validate the source comparing the results of the home locations detected by Twitter and the census. The results show a strong correlation, indicating that the methodology is appropriate and that the possibilities of this information are vast.

Keywords: Big Data; Mobility; Geographical Information Systems

1. INTRODUCCIÓN

El acceso a nuevas fuentes de información, sean de tipo colaborativo, aportadas por el usuario, o derivadas de la gestión inteligente de las ciudades, combinadas con las herramientas TIG y la cartografía permite abordar nuevos análisis que proporcionan una información que en ocasiones suplementa las limitaciones de algunas de las fuentes tradicionales. Sin embargo, exige una depuración de los datos, un mecanismo de control de resultados y un tratamiento previo en combinación con nuevos marcos teóricos y estrategias metodológicas para poder extraer todas sus potencialidades. En esta comunicación pretendemos realizar una aportación en esta línea de trabajo proponiendo un proceso de filtrado de datos procedentes de fuentes sociales, en este

caso Twitter, para analizar patrones de movilidad de la población en un AM. Se elige dicha plataforma porque permite descargar información geolocalizada de sus usuarios y por ser una de las redes sociales más utilizada.

El interés de este trabajo es amplio, ya que aúna dos campos de estudio, la movilidad metropolitana y el análisis de las posibilidades que ofrecen los Big Data. Tradicionalmente la generación de la información para un estudio de movilidad ha sido una parte compleja y costosa, dependiente de trabajos específicos (realización de aforos, encuestas...). Actualmente el uso de redes sociales posibilita recabar información en tiempo real y sin coste específico para la investigación, aunque partiendo de la base de que los datos se han producido con otro objetivo, por lo que se requerirá un amplio esfuerzo de detección y extracción de aquellos válidos para el fin propuesto. Por otro lado, el estudio de las posibilidades del Big Data en sí mismo es un campo novedoso y con gran interés para la Geografía. Se trata de información hasta ahora inexistente, donde millones de usuarios generan rastros de sus actividades cotidianas que pueden contener información sobre diversas esferas de su vida. El valor de esta información puede ser excepcional bajo una correcta gestión y es específicamente ahí donde se enmarca la finalidad de la presente comunicación.

El trabajo que se presenta forma parte de un estudio más amplio que se está llevando a cabo en la actualidad, por lo que son muchas las líneas de trabajo abiertas que se están desarrollando. En la presente comunicación nos centramos en el conocimiento de la fuente y las características de los datos que contiene, pudiendo así determinar su validez para según qué trabajos, así como en la identificación del lugar de residencia de la persona que emite los mensajes, con el fin de realizar “a posteriori” un seguimiento de sus desplazamientos en base al lugar de emisión.

De forma específica, los objetivos principales del trabajo son:

1. Conocer la naturaleza de los datos geolocalizados que provienen de Twitter
2. Elaborar una propuesta metodológica para el filtrado de los datos con el fin de capturar patrones de movilidad
3. Analizar la validez de esta fuente para capturar dichos patrones, centrándonos en la detección de su lugar de residencia como punto base.

El trabajo se presenta empleando una estructura convencional, partiendo de la revisión del estado de la cuestión que ofrece diferentes enfoques y la detección de posibles problemáticas a tratar. A partir de esta revisión y de un trabajo de análisis específico se expone la metodología propuesta, sobre la que se trabaja con la muestra de datos obtenida. El análisis de los resultados para la provincia de Valencia ofrece una visión clara de la validez de la fuente y del uso que se puede hacer de la misma.

2. ESTADO DE LA CUESTIÓN

La revisión de trabajos anteriores es la base para plantear el trabajo actual, posibilitando un estudio fundamentado y comparado. En este apartado se señalarán únicamente los trabajos que se emplean como referentes para definir la metodología propuesta.

Uno de los trabajos de referencia es el de Haung et al (2015), en donde se analizan las pautas de movilidad humana empleando una perspectiva individual espaciotemporal a partir de datos Twitter. En ese trabajo, la recopilación de datos durante un amplio periodo de tiempo les permitió obtener la evolución temporal de las posiciones individuales, reconociendo patrones regulares e irregulares de desplazamiento. La forma de abordar este análisis de movilidad se basa en los estudios de Kwan (1999), quien afirmó que las pautas comunes de movilidad pueden reconocerse a través de múltiples trayectorias individuales. Sin embargo, el

estudio de Haung antes citado basa su filtrado de datos en seleccionar usuarios con gran número de tweets geolocalizados y en eliminar manualmente aquellos que corresponden a organizaciones, lo que no se ajusta estrictamente al trabajo que aquí desea desarrollarse.

Más interesante desde nuestra perspectiva es el trabajo de Jurdak et al (2015), cuyos resultados concluyen la existencia de correlación entre el lugar desde donde más tweets realiza un usuario, que denominaremos posición dominante, y su residencia, lo que constituye la hipótesis de partida de la presente comunicación. En ese trabajo se emplean diversos procedimientos de filtrado de datos orientados a conocer y eliminar sesgos provenientes de la información. En primer lugar, a partir de distintas funciones, definen una distancia de 100 m como umbral máximo para determinar movimientos realizados dentro lo que consideran un mismo lugar. Por otra parte, partiendo de la idea de que los 140 caracteres que contiene un tweet puedan ser escasos para que un usuario se exprese, consideran la posibilidad de que algunos de ellos realicen tweets consecutivos desde una misma ubicación, lo que sesgará el cálculo de posiciones dominantes. Esta posibilidad se incrementa ante ciertos eventos que aumentan el uso de Twitter en un momento determinado (deportivos, políticos, sociales...). Para eliminar el impacto de estas actitudes se integrará un filtrado específico.

En el presente trabajo se adopta la perspectiva de obtener pautas generales a partir de las individuales de Kwan (1999) y Haung et al. (2015), desarrollando la hipótesis de Jurdak et al. (2015) para la identificación del lugar de residencia del usuario. Por otra parte, tomamos el límite de los 100m definido por Jurdak et al. (2015) como distancia mínima para considerar que se ha producido movimiento. Por el contrario, a diferencia de los autores mencionados, la selección de usuarios se realizará teniendo en cuenta cantidad de posiciones diferentes, no un número mínimo de tweets. Finalmente, de los trabajos de Hasan (2013) y Gabrielli et al (2014), que combinan datos de Twitter y Foursquare para caracterizar los desplazamientos urbanos en función a la finalidad de los mismos, recogemos la idea de la importancia de diferenciar las tipologías de las posiciones observadas. Es decir, que la fuente de datos contiene información de diversas actividades y es necesario reconocer el motivo del desplazamiento para obtener buenos resultados.

3. ÁMBITO DE ESTUDIO

La presente investigación se enmarca en el estudio de la movilidad en el ámbito del Área Metropolitana de Valencia, si bien en esta primera aproximación se han tomado datos referentes a la provincia de Valencia. Por un lado, de esta forma la cantidad de municipios testeados aumenta y permite ver qué ocurre en las zonas más rurales con menor población. Por otro lado, tendremos así la posibilidad de trabajar con las distintas delimitaciones del AMV, observando si los patrones de movilidad marcados por Twitter responden mejor a unas delimitaciones u otras, e incluso dejando abierta la posibilidad de trazar un área nueva. La dimensión del área de estudio condiciona los resultados, ya que su tamaño es mediano en relación a los ámbitos tratados en estudios de otros autores, por lo que la generación de datos de este tipo es menor a la de otros entornos analizados desde esta perspectiva. La recogida de datos se enmarca entre las fechas 5/06/2015 y 21/02/2016, habiendo sido posible recabar información de 146 días, 22.641 usuarios y un total de 177.675 tweets.

4. METODOLOGÍA

El punto de partida es la programación de la API de Twitter para la descarga en tiempo real de la información georreferenciada que se produce en el área seleccionada. Tras el acopio de un periodo de tiempo amplio (146 días), se procede a realizar este primer análisis de datos empleando como herramienta fundamental los Sistemas de Información Geográfica (SIG).

Debido a que el trabajo es en sí un análisis de la fuente, será más representativo para su exposición explicar el procedimiento de filtrado conforme se vaya conociendo la fuente, por lo que en este punto se presenta un esquema básico de la metodología y será en el apartado de resultados donde se encuentre en detalle el

proceso. La propuesta metodológica para el uso de Twitter como fuente de datos para capturar patrones de movilidad es la siguiente:

1. Estudio de la fuente:
 - a. Datos contenidos
 - b. Generación de la información y precisión espacial asociada
2. Filtrado de los datos para obtener información de movilidad:
 - a. Filtrado de usuarios por fuente: Se eliminan los datos que no han sido generados por personas individuales
 - b. Filtrado de tweets por comportamiento: Se eliminan los tweets masivos
 - c. Filtrado de usuarios según movilidad reflejada: Se eliminan los usuarios con movilidad poco representativa.
3. Detección del lugar de residencia y validación con el padrón.
4. Conclusiones

5. DESCRIPCIÓN Y ANÁLISIS DE RESULTADOS

5.1. Estudio de la fuente:

a. Datos contenidos:

Los datos procedentes de la API de Twitter que se pueden descargar para cada tweet georreferenciado y que utilizaremos para estudiar la fuente son los siguientes:

- Latitud, Longitud (Datum WGS84)
- Fecha y hora de producción del Tweet
- Identificador único del usuario
- Nombre de usuario
- Localización: seleccionada por el usuario, no tiene por qué corresponderse con una localización real.
- Idioma: seleccionado por el usuario, no tiene por qué corresponderse al utilizado en la generación de tweets.
- Fuente del tweet: Twitter, Instagram, Foursquare, Endomondo, etc. Permite conocer el tipo de datos que comparte y además constituye la base para conocer la forma en que se generan las posiciones y su precisión espacial.
- Identificador único del Tweet
- Texto del Tweet

A partir de estos datos se crean las unidades puntuales que localizan en el tiempo y el espacio cada tweet y su información asociada.

b. Cómo se generan los puntos de posición

La primera característica a tener en cuenta es que la información georreferenciada de la que hacemos uso no proviene únicamente de la propia plataforma de Twitter, sino que son muchas otras las que la utilizan para enlazar sus propios datos. Es decir, que otras redes sociales tienen la opción de conectar con la cuenta Twitter

del usuario para volcar también sus contenidos. En nuestro caso de estudio, las fuentes de datos con más de 100 tweets registrados se distribuyen de la siguiente forma:

Fuente	Tweets	Porcentaje	Válidas Movilidad
Instagram	80.254	45,39%	si
Twitter	65.368	36,97%	si
Foursquare	7.725	4,37%	si
Tráfico	6.341	3,59%	no
Tendencias	6.089	3,44%	no
Tiempo	5.391	3,05%	no
Deporte	1.844	1,04%	si
Plataforma externa	1.795	1,02%	si
Foros de localidades	1.411	0,80%	no
Hoteles, atracciones...	361	0,20%	si
Juegos	224	0,13%	no

Tabla 1. Fuente inicial de los tweets. Fuente: Elaboración propia a partir de datos de la API Twitter

A través la tabla 1 es posible reconocer las principales fuentes que vuelcan información en Twitter; de ellas, las tres principales, Instagram, Twitter y Foursquare, contienen el 86,73% del total. De esto se deriva también que el 63,03% de los tweets provienen inicialmente de plataformas distintas a Twitter. La importancia de este dato es que cada una de ellas georreferencian la información bajo sus propios parámetros, por lo que se debe conocer cuál es su funcionamiento básico. Haremos un recorrido por las principales:

En **Twitter** hay dos maneras de usar servicios de ubicación:

- Pulsando el marcador de “localización” al redactar el tweet y seleccionando de forma manual a partir del listado que ofrecen la localización que se desea etiquetar
- Activando el botón específico de “Compartir localización exacta”, de forma que la aplicación accede al GPS y comparte la latitud y longitud con la precisión del dispositivo GPS del móvil.

La API no incluye información del procedimiento por el que se ha geolocalizado el tweet, por lo que tenemos datos de distinta precisión que provienen de la misma fuente.

En **Instagram**, plataforma que produce más tweets geolocalizados que el propio Twitter, los servicios de ubicación funcionan como la primera opción de Twitter, de forma que es el usuario quien selecciona su posición a partir de un listado específico. Este listado no contiene una gran base de datos con el callejero, sino que principalmente contiene municipios, barrios, comercios o lugares emblemáticos. Se ha eliminado la opción directa de crear un lugar nuevo desde Instagram, debiendo emplear otras plataformas para ello y dificultando así este trabajo. Esto hace que las posiciones reveladas desde Instagram sean más generales, ya que no permite usar la localización GPS del usuario ni contiene un buen callejero. En consecuencia, será complicado establecer lugares de residencia con precisión a través de tweets generados desde esta plataforma.

La utilización de **Foursquare** es similar a Instagram, pero con sus propios listados de direcciones. Es decir, se seleccionan lugares específicos, pero sus coordenadas no coinciden con las de Instagram ni con las de Twitter.

Como se ha visto, la precisión de los datos es distinta según su procedencia y, sobre todo, según la forma de interactuar de cada usuario. Esto supone que la precisión del GPS de los dispositivos no sea crucial para este trabajo, pero sí lo sea reconocer el tipo de información que cada usuario decide mostrar de sí mismo. Es decir, un usuario que crea un tweet geolocalizado pero que selecciona como opción de localización en su listado

“Comunidad Valenciana” o “Valencia” no muestra una información válida para el estudio. Inicialmente se ha obviado este hecho, con el fin de reconocer si tiene impacto o no tras los filtrados inicialmente propuestos. Como se podrá ver en los resultados, sí lo tiene y debe ser objeto de tratamiento específico en este tipo de trabajos.

5.2. Filtrado de los datos para obtener patrones de movilidad

a. Filtrado por fuente

Twitter es una plataforma que se utiliza para múltiples propósitos, entre ellos ofrecer información del estado del tráfico, meteorología, tendencias, foros, etc. Las posiciones que revelan este tipo de datos, la mayoría provenientes de sensores, no encierran información sobre la movilidad de los ciudadanos, por lo que deben ser eliminadas. En la Tabla 1 se ha realizado una primera agrupación de fuentes en función del tipo de datos y de si éstos son *a priori* válidos para representar la movilidad humana (3ª columna). Esta información será objeto del primer filtrado a realizar, cuyo resultado se encuentra en la Tabla 2. Aquí vemos que 14 usuarios han generado 19.456 tweets no válidos.

b. Filtrado por comportamiento:

Como se ha comentado, el comportamiento que tenga un usuario dentro de la red social varía en función a ciertos eventos de su vida, que pueden corresponderse o no con eventos generales. En consecuencia, si en un momento dado un usuario realiza muchos tweets desde una misma posición, podemos obtener una posición dominante afectada por este momento puntual. Para eliminar el efecto de las actitudes planteadas, se filtrarán los tweets generados en un intervalo menor a tres horas, que puede relacionarse con un evento específico, y en una distancia de 100m. Tras analizar los resultados de este paso, se ha decidido modificar el máximo de tres horas por un día completo, ya que se observa que hay usuarios mucho más compulsivos de lo esperado. El resultado se encuentra en la Tabla 2, donde se puede observar que el número de tweets se reduce en un 35%, mientras el número de usuarios se mantiene, como era de esperar.

c. Filtrado por movilidad reflejada

En este punto se eliminan los usuarios con movilidad poco representativa. Para ello se hace una diferenciación entre la cantidad de posiciones diferentes que muestra cada usuario y el periodo de tiempo sobre el que han hecho uso de la plataforma. Dado que en este caso interesa conocer la movilidad de la población residente, se eliminarán los usuarios que puedan corresponder a turistas utilizando las fechas de sus tweets. En este tercer filtrado se quedarían fuera también negocios o instituciones que se publiciten en Twitter, ya que estos lugares suelen tener su propia dirección en el listado, por lo que los puntos se localizarán siempre en el mismo lugar. En la tabla 2 se recogen los dos filtrados anteriores y las diferentes posibilidades del tercer filtrado, con el fin de observar cómo responden los datos a las restricciones de posiciones y tiempo aplicadas:

Filtrado	Tweets	Usuarios
0. Ninguno	177.675	22.641
1. No humanos por fuente	158.219	22.627
2. Comportamiento	103.518	22.627
3. Sin movilidad o poco representativa		
3.1. Sólo 1 posición	13.105	11.423
3.2. Entre 2 y 4 posiciones	21.824	6.905
3.3. Más de 5 posiciones	68.589	4.299
3.3.1. Diferencia >= 15 días	66.722	3.882
3.3.2. Diferencia >= 60 días	56.378	3.007

Tabla 2. Filtrado de tweets. Fuente: Elaboración propia a partir de datos de la API Twitter

El desglose de los datos para realizar el tercer filtrado muestra características muy interesantes de la fuente, como por ejemplo que los usuarios que sólo tienen una posición registrada suponen el 50,5% del total (Punto 3.1). Aquí se agrupan principalmente los usuarios esporádicos de la plataforma y aquellos que twitteen desde un negocio o institución. Se considera que un usuario que revela un mínimo de 5 posiciones diferentes es un usuario avanzado que permite realizar un buen seguimiento de sus rutinas y sobre el que se puede localizar con mayor fiabilidad su lugar de residencia, por lo que se tomará como primera restricción. A su vez, dentro de los usuarios que cumplen esta condición, y con el fin de eliminar los no residentes habituales que realizan una estancia puntual en el área, se aplica la restricción de haber realizado tweets separados en el tiempo por un mínimo de 60 días. Por tanto, al final del proceso se cuenta con 3.007 usuarios adecuados para calcular su lugar de residencia, partiendo de la suposición que la posición dominante desde la que twitteen es su residencia, como ya se ha comentado.

5.3. Obtención de resultados

Como ya se ha mencionado, para conocer el lugar de residencia de cada usuario se parte del supuesto de que se corresponde con el lugar desde el que se emplea más la plataforma, por lo que se calcula la posición dominante de cada uno de ellos. Este cálculo se hace empleando el radio de 100 metros propuesto por Jurdak (2015) y ya empleado anteriormente para el filtrado de tweets compulsivos. Se selecciona la posición más repetida por usuario, que debe ser mayor a 1. El cálculo muestra que de los 3.007 usuarios 362 nunca repiten ubicación, por lo que no será posible realizar esta asignación. Obtenemos así un total de 2.645 usuarios, que denominaremos “trazables”, a los que se les ha asignado un lugar de residencia.

Para validar esta información se realiza un Test de correlación de Pearson que compara la cantidad de usuarios localizados en cada municipio a través de Twitter y el número de habitantes según los datos del padrón municipal 2015. Se utiliza la escala municipal salvo en el caso del municipio de Valencia, que, debido a su tamaño demográfico, se divide en distritos.

El resultado inicial ofrece un coeficiente de correlación $R= 0,72$, lo cual quiere decir que el padrón y los residentes detectados por Twitter se encuentran bien correlacionados. No obstante, el análisis del diagrama de dispersión (figura 1) muestra tres puntos que no siguen el patrón marcado por los demás. Estos tres puntos son los dos distritos centrales del municipio de Valencia y el distrito del puerto del mismo municipio.

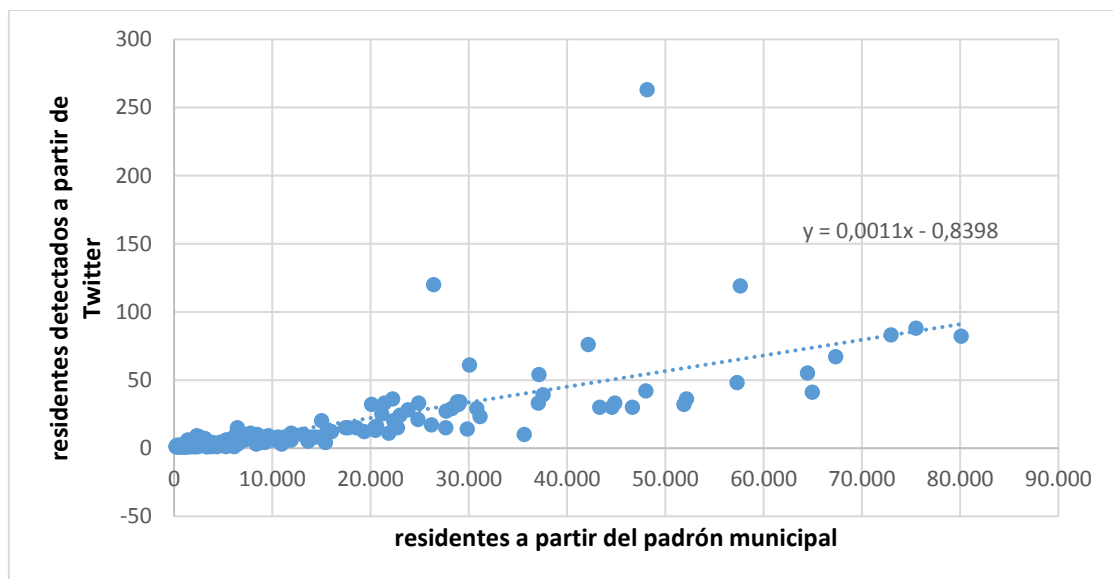


Figura 1. Resultado primera detección de residencias y comparación con el padrón 2015

Un análisis de detalle de estos tres distritos atípicos revela el impacto del uso de las localizaciones generales que describíamos en el punto 5.1 b). En el caso de los distritos del centro, se detecta que Instagram, Twitter y Foursquare colocan aquí a los usuarios que marcan “Valencia” en su listado de lugares, una localización general que no tiene la precisión suficiente para hacer una detección de lugar de residencia o un análisis de movilidad, lo cual induce a error. En el caso de la zona del puerto, la explicación parece estar en el fuerte impacto que tiene la localización de lugares de ocio nocturno y la realización de eventos en la emisión de tweets. Este resultado muestra la necesidad de filtrar las posiciones sin precisión espacial suficiente y de detectar eventos que puedan ser tenidos en cuenta para analizar la pauta de movilidad de un usuario, pero no para localizar su residencia. Con ello, se realiza un filtrado final de estos puntos, obteniendo así un resultado en el Test de Pearson de una adecuación $R= 0,88$, lo que indica una fuerte correlación de variables (ver figura 2). Los resultados no muestran sesgos importantes para los distintos grupos de población por edades y sexo, tal y como se muestra en la tabla 3, aunque sí un coeficiente de correlación ligeramente superior en el caso de la población entre 25 y 34 años.

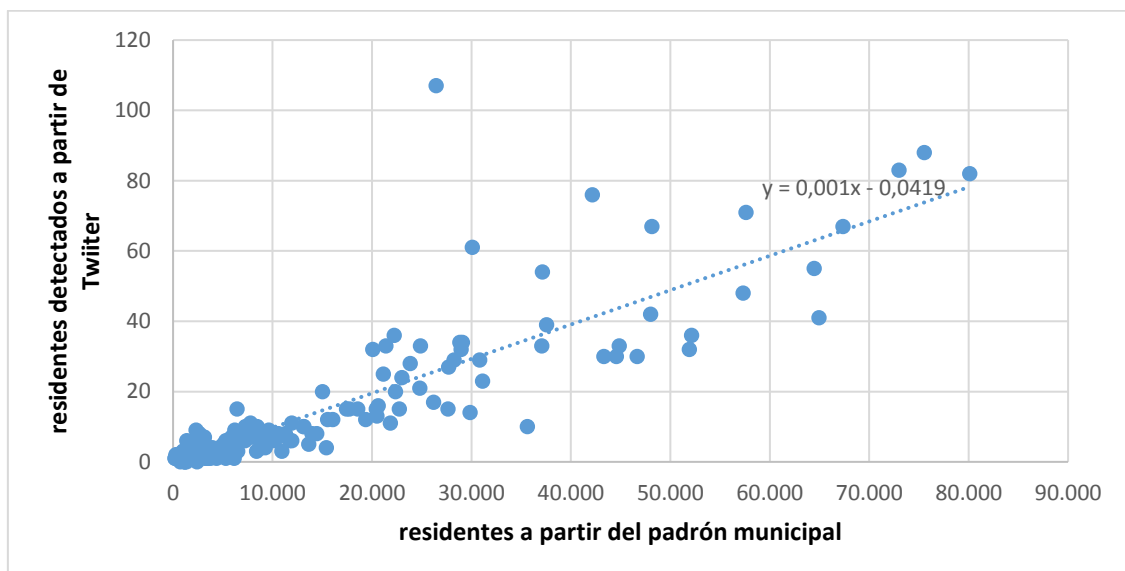


Figura 2. Resultado detección de residencias y comparación con el padrón 2015

Población 2015	Coficiente R
Hombres	0,874670
Entre 15 y 24 años	0,865535
Entre 25 y 34	0,879550
Entre 35 y 44	0,865312
Mujeres	0,883120
Entre 15 y 24 años	0,871964
Entre 25 y 34	0,884486
Entre 35 y 44	0,867366
Total Población	0,879374

Tabla 3. Coeficiente de correlación entre el número de localizaciones residenciales y distintos grupos de población

Estos datos corroboran que la tendencia de los usuarios a usar Twitter desde su lugar de residencia es alta, según indicó Jurdak (2015). Pero no todos los usuarios siguen este patrón, y en algunos casos la posición dominante se ha detectado en lugares que se corresponden con espacios de ocio. Los efectos se observan

tanto en la figura 2, en donde el punto que no sigue la pauta marcada por los demás corresponde al distrito centro del Ayuntamiento y barrio del Carmen, lugares habituales para realizar actividades de ocio; como en la figura 3, en donde los mayores valores de residuales positivos corresponden a los espacios urbanos, en particular aquéllos con importantes actividades de ocio, tales como Gandía o Cullera. Este dato nos recuerda el origen social de los datos con los que se trabaja y la clara influencia de estas actividades.

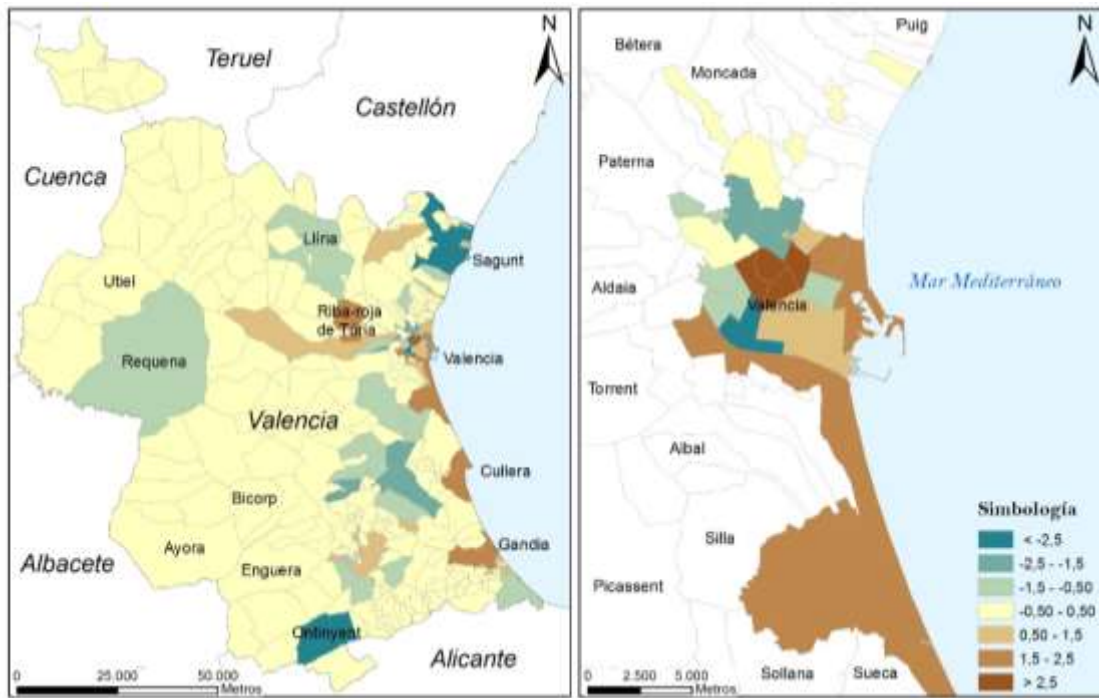


Figura 3. Residuales estandarizados de la regresión entre número de habitantes en 2015 y número de residencias detectadas. Fuente: INE y elaboración propia

6. CONCLUSIONES

Los objetivos planteados en el trabajo se han desarrollado de forma satisfactoria. El estudio de la naturaleza de los datos geolocalizados de Twitter ha permitido plantear una metodología válida para su filtrado y para el cálculo de los lugares dominantes asociados con lugares de residencia. Dicho estudio nos ha posibilitado comprender las desviaciones y los datos que modificaban el modelo erróneamente, marcando requerimientos de filtrado no propuestos inicialmente en la metodología. Se ha obtenido también una orientación sobre la posibilidad de mejorar la detección de lugares residenciales empleando nuevos parámetros como momentos temporales o palabras clave. En la actualidad se siguen recabando datos que nos ofrecerán una muestra mayor y posibilitarán refinar el modelo y obtener una mayor significación.

7. AGRADECIMIENTOS

Esta comunicación se ha elaborado en el marco del proyecto de investigación "Sostenibilidad y competitividad urbanas en un contexto global. El Área Metropolitana de Valencia" (CSO2013-46863-C3-1-R) financiado por el Ministerio de Economía y Competitividad dentro del Programa Estatal de Investigación, Desarrollo e Innovación Orientada a los Retos de la Sociedad, modalidad 1, "Retos Investigación": Proyectos de I+D+I; Carmen Zornoza cuenta con una ayuda para contratos predoctorales para la formación de doctores (BES-2014-067846) dentro del Subprograma Estatal de Formación del Programa Estatal de Promoción del Talento y su Empleabilidad, en el marco del Plan Estatal de Investigación Científica y Técnica y de Innovación 2013-2016, financiado por el Ministerio de Economía y Competitividad y cofinanciado por el Fondo Social Europeo.

8. BIBLIOGRAFÍA

Gabrielli, L., Rinzivillo, S., Ronzano, F., & Villatoro, D. (2014): "From Tweets to Semantic Trajectories: Mining Anomalous Urban Mobility Patterns", Jordi Nin, Daniel Villatoro (Ed.): *"Citizen in Sensor Networks"*. Switzerland, Springer International Publishing. pp.26-35

Hasan, S., Zhan, X., & Ukkusuri, S. V. (2013): "Understanding urban human activity and mobility patterns using large-scale location-based data from online social media". *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*, Agosto 2013, Chicago, USA.

Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M., & Newth, D. (2015): "Understanding human mobility from Twitter". *PloS one*, 10(7), e0131469.

Kwan, M. P. (1999): "Gender, the home-work link, and space-time patterns of nonemployment activities". *Economic geography*, 75(4), pp. 370-394.

Song, C., Qu, Z., Blumm, N., & Barabási, A. L. (2010): "Limits of predictability in human mobility". *Science*, 327(5968), pp. 1018-1021.