



Gestión de datos masivos

Apellidos, nombre	Llopis Castelló, David ¹ (dalocas@upv.es) Camacho Torregrosa, Francisco Javier ¹ (fracator@tra.upv.es)
Departamento	¹ Ingeniería e Infraestructura de los Transportes
Centro	Universitat Politècnica de València

1 Resumen de las ideas clave

En este artículo se presentan las diferentes tipologías de sistemas y dispositivos de recogida de datos masivos para conocer y analizar principalmente la movilidad de los ciudadanos tanto en entornos urbanos como interurbanos.

2 Objetivos

Una vez que el estudiante haya leído con detenimiento este documento, será capaz de:

- Definir los distintos sistemas y dispositivos que nos permiten recoger información sobre la movilidad tanto en entorno urbano como interurbano.
- Identificar las ventajas y desventajas del empleo de cada uno de estos dispositivos.
- Aplicar técnicas básicas para el tratamiento de los datos almacenados.

3 Introducción

Hoy en día el estudio de la movilidad para conocer y entender el comportamiento de las personas se ha convertido en uno de los grandes retos relacionados con la ingeniería de tráfico tanto en entornos urbanos como interurbanos.

La concentración de las personas en las ciudades ha provocado un aumento significativo del tráfico motorizado urbano, lo que ha conducido a un aumento de la congestión, una mayor concentración de gases de efecto invernadero en el aire que respiramos y un incremento del nivel sonoro. Por ello, muchas de las políticas propuestas por las administraciones locales es fomentar el uso de vehículos de movilidad personal, tales como la bici y el patín eléctrico, para descongestionar la ciudad y hacerla más sostenible medioambientalmente.

Asimismo, la movilidad en entorno interurbano también ha experimentado un aumento en los últimos años debido, principalmente, a la mejora de la funcionalidad y el aumento de la capacidad de la red de carreteras, así como al auge económico tras la crisis que comenzó en el año 2008 (Figura 1). Así pues, es fundamental conocer los patrones de comportamiento de los usuarios con el fin de determinar qué infraestructuras pueden estar más necesitadas.

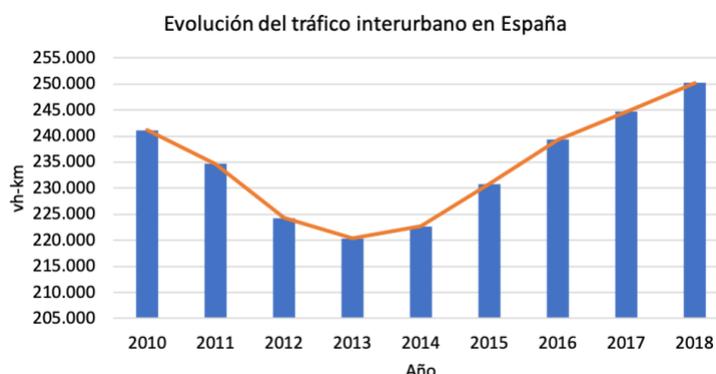


Figura 1. Evolución del tráfico interurbano.



Antes de conocer los distintos tipos de sistemas y dispositivos que nos permiten recoger información sobre la movilidad de los ciudadanos, ¿cuál es tu opinión sobre la aplicación de la inteligencia artificial en la ingeniería de tráfico?

Para ello, visualiza los siguiente vídeos:

<https://www.youtube.com/watch?v=Rg9PN6DSAfc>

<https://www.youtube.com/watch?v=dS4pWnNlxfA>

4 Encuestas

Una encuesta consta de una serie de preguntas que se hace a un número significativo de personas para reunir datos o para detectar la opinión pública sobre un asunto determinado. Una de sus principales ventajas es la gran capacidad que tiene para conseguir información muy detallada sobre patrones de movilidad de las personas, así como características sociodemográficas. No obstante, las encuestas presentan grandes limitaciones, destacando:

- Requieren de recursos económicos importantes: para obtener una muestra que represente adecuadamente los patrones de movilidad de la población necesitamos realizar un enorme número de encuestas que, generalmente, supera el número de encuestas que podemos conseguir en un tiempo razonable.
- Plazo de ejecución limitado: la realización de encuestas puede llevar desde pocas semanas a meses e incluso años. En este sentido, es importante definir bien el objetivo de la recogida de datos porque podría ser necesario que la encuesta se tuviese que llevar a cabo en un reducido plazo de tiempo. Por ejemplo, si queremos conocer profundamente los desplazamientos que los ciudadanos realizan a lo largo del día durante un determinado mes del año, no podemos dilatarlos mucho en el tiempo porque la veracidad de los datos dependerá en gran medida del tiempo transcurrido entre el día del que queremos disponer de datos y la realización de la encuesta.
- Precisión y veracidad de las respuestas: a la hora de realizar una encuesta debemos tener en cuenta que muchos de los encuestados no tienen o tienen poco interés en rellenarla. Por ello, los cuestionarios largos (> 5-10 min) suelen conducir a encuestas incompletas y con datos anómalos. Por tanto, nuestras encuestas deberán ser breves y utilizar un lenguaje sencillo y preciso.
- Planteamiento de la encuesta: antes de lanzar la encuesta de manera masiva es interesante realizar una prueba piloto con el fin de acabar de perfilar la encuesta. Con ello se pretende identificar si las respuestas que obtenemos de los usuarios son válidas para estudiar lo que realmente queremos investigar o analizar. En este punto, debemos tener en cuenta de que si omitimos en nuestra encuesta un factor relevante de nuestro estudio, entonces la encuesta no habrá servido para nada y habremos perdido importantes recursos desde el punto de vista económico y temporal.

A continuación, se presentan diferentes tipos de encuestas destinadas a conocer la movilidad de las personas durante un periodo de tiempo determinado.

4.1 Encuestas domiciliarias

Se llevan a cabo en el domicilio de las personas encuestadas, permitiendo obtener datos de todos los ocupantes de la vivienda. Dentro de este tipo de encuestas, debemos conocer las encuestas de diario de viajes, que tratan de que la persona encuestado recoja en un diario todos los movimientos que realiza en un periodo de tiempo específico, que posteriormente debe entregarnos. Este tipo de encuestas nos permite obtener información muy detallada.

4.2 Encuestas de interceptación

Tratan de interceptar al usuario en un punto determinado de la red de transporte como, por ejemplo, carreteras de acceso o salida del área de estudio, estaciones de medios de transporte público o puntos de intercambio modal. Generalmente las utilizamos para recoger datos de usuarios no residentes (personas que no podemos encuestar en su domicilio) y también para ampliar y corroborar los datos de las encuestas domiciliarias. Especialmente este tipo de encuesta debe ser breve para no causar excesivas molestias al encuestado y basarse en información básica del viaje (origen/destino/propósito).

Finalmente, es interesante conocer dos tipos de encuestas de interceptación: (i) cordón externo y (ii) cordón interno (Figura 2). Las encuestas de cordón externo nos permiten conocer el comportamiento de los usuarios que entran y salen de un área urbana específica, mientras que las encuestas de cordón interno tienen como objetivo determinar los movimientos que llevan a cabo los usuarios dentro de un área urbana determinada. Cabe destacar que las encuestas de cordón externo habitualmente requieren de menos medios que las de cordón interno porque los accesos a las ciudades normalmente están muy bien definidos y, por tanto, se necesita de menos recursos humanos para realizar la encuesta.



Figura 2. Encuestas de interceptación (cordón externo: rojo, cordón interno: amarillo).

4.3 Encuestas telefónicas

Las encuestas telefónicas se basan en realizar llamadas telefónicas a una muestra de personas y recopilar datos a través de cuestionarios previamente elaborados. Entre sus ventajas más interesantes podemos destacar el reducido coste, ahorro en tiempo de ejecución, la diversidad de los encuestados y el feedback con el encuestado, pues la



comunicación es directa a través de un medio impersonal. No obstante, también poseen ciertas desventajas, como la baja tasa de respuestas, restricciones en el diseño del cuestionario (breve y sin posibilidad de apoyarse en elementos visuales), poca fiabilidad de las respuestas y, en ocasiones, requiere de un contacto previo con el encuestado.

4.4 Encuestas online

Hoy en día, las encuestas en línea se han vuelto tan populares, que podría decirse que son el recurso número uno utilizado para la recolección de datos en estudios e investigaciones. Este tipo de encuestas permite obtener datos de un gran número de entrevistados, requieren de poco personal y mínimos recursos económicos, presentan una notable mejoría con respecto al resto de tipos de encuestas en cuanto al método de recopilación de información y permiten ejercer un mayor control sobre los encuestados, pues podemos conocer el tiempo que tarda el encuestado en responder cada una de las preguntas. Sin embargo, sus principales desventajas tienen que ver con el problema para localizar muestras representativas y un incremento de trabajo por parte del entrevistado debido a que no cuentan con la presencia de un encuestador.

5 Sensores de tráfico interurbano

Los sensores utilizados para identificar la presencia de tráfico pueden ser clasificados en base a la ubicación de los mismos y la necesidad o no de que los vehículos dispongan de ciertos dispositivos. De este modo, podemos diferenciar entre:

- Sensores de tráfico autónomos: el sensor está ubicado en la infraestructura y no se requiere de ningún dispositivo embarcado en los vehículos.
- Sensores de tráfico dependientes: el sensor está en la infraestructura y requiere que el vehículo disponga de un dispositivo determinado o esté equipado con sensores.

En las siguientes secciones aprenderás un poco más sobre estos tipos de sensores.

5.1 Sensores de tráfico autónomos

Principalmente podemos distinguir dos tipos de sensores de tráfico autónomo: (i) intrusivos y (ii) no intrusivos. Mientras que las tecnologías intrusivas están instaladas en o a lo largo del pavimento de la carretera, las tecnologías no intrusivas se encuentran por encima o a los lados de la carretera causando el mínimo efecto sobre el flujo de tráfico.

5.1.1 Sensores intrusivos

Este tipo de sensores pueden proporcionar información sobre el volumen de tráfico, detección y clasificación de vehículos, e incluso información sobre la velocidad. La principal ventaja de los sensores intrusivos en comparación con los no intrusivos es su reducido coste. No obstante, presentan importantes inconvenientes como, por ejemplo, la interrupción del tráfico durante la instalación y la reparación, conteos erróneos en caso de que la calzada no sea apropiada o por una mala instalación y la necesidad de reinstalación en caso de operaciones de mantenimiento y rehabilitación.

Entre la multitud de sensores intrusivos existentes, cabe destacar que los más empleados en carreteras son las espiras magnéticas y los tubos neumáticos. Las espiras magnéticas son

los sensores más extendidos en las carreteras españolas ya que se trata de una tecnología barata y muy desarrollada que no se ve afectado por las condiciones ambientales (Figura 3a). Además, su funcionamiento basado en la variación de la impedancia que se registra en la espira al paso de un vehículo es muy simple.

Por otro lado, los tubos neumáticos son tubos de goma dispuestos de manera transversal a la calzada, de modo que son capaces de detectar el paso de vehículos a partir del cambio de presión sobre el aire contenido en el tubo (Figura 3b). La ráfaga de aire generada dentro del tubo excita una membrana que cierra un interruptor produciendo una señal eléctrica que se transmite a un contador o a un equipo con software apropiado para su procesamiento.



(a) Espiras magnéticas.



(b) Tubos neumáticos.

Figura 3. Sensores intrusivos

5.1.2 Sensores no intrusivos

A diferencia de los sensores intrusivos, este tipo de detectores no interfieren directamente sobre el tráfico durante su funcionamiento puesto que se ubican sobre o en el margen de la carretera. Los tipos de detectores no intrusivos empleados actualmente son dos:

- Sensores activos: emiten una señal y captan la respuesta reflejada sobre el vehículo (radares de microondas, radares láser y los sensores ultrasónicos).
- Sensores pasivos: captan variaciones producidas, en ciertos parámetros, por el paso de un vehículo (cámaras de vídeo, sensores infrarrojos y sensores acústicos).

Entre todos los sensores cabe destacar los radares de microondas y láser y las cámaras de vídeo. Concretamente, los radares de microondas son los empleados por los controles fijos de velocidad (Figura 5a). Estos emiten una onda continua y detectan el paso de los vehículos al mismo tiempo que miden su velocidad y longitud. Además, puede detectar el paso de los vehículos por varios carriles y no se ve afectado por las condiciones meteorológicas.

En cuanto al radar láser, este es el empleado por la Guardia Civil de tráfico para el control de la velocidad en ubicaciones variables (Figura 5b). Este dispositivo emite varios haces de luz que permite obtener de una manera muy precisa el tipo de vehículo, su velocidad y su posición. No obstante, el funcionamiento del sensor puede verse afectado por la niebla, siempre que la visibilidad sea inferior a 6 metros.

Por último, las cámaras de vídeo son un elemento pasivo que requiere de algoritmos de visión artificial basados habitualmente en técnicas de aprendizaje profundo (*Deep Learning*) para la identificación y clasificación de los vehículos (Figura 5c). Adicionalmente, puede ofrecer datos de la velocidad y lectura de matrículas, lo que permite realizar aforos y controles de la velocidad en tramos. A pesar de que los algoritmos que emplea este tipo de sensor son cada vez más fiables, el funcionamiento de las cámaras está altamente influenciado por las condiciones meteorológicas y lumínicas.

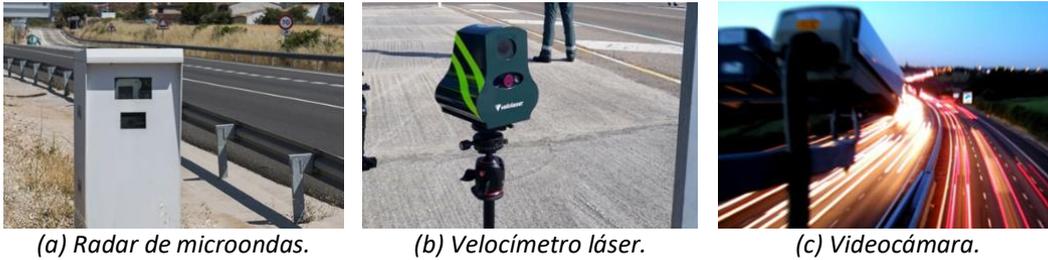


Figura 4. Sensores no intrusivos.



Ahora es el momento de identificar el formato de los datos recogidos por un velocímetro láser y filtrar dicha información. Para ello, visualiza el siguiente vídeo.

<https://media.upv.es/player/?id=a74787c0-69c8-11ea-ab9b-6df6378bebd6>

Si además quieres practicar, el archivo de datos lo puedes descargar haciendo click [aquí](#).

5.2 Sensores de tráfico dependientes

A diferencia de los sensores de tráfico independientes, los denominados sensores de tráfico dependientes requieren la instalación o presencia de un dispositivo en los vehículos. De esta manera, los sensores ubicados en la infraestructura son capaces de identificar la posición de cada vehículo a partir de la detección de alguno de los dispositivos dispuestos en su interior en base, habitualmente, a dos tecnologías:

- Radiofrecuencia
- Sistemas de navegación por satélite o GNSS

La identificación por radiofrecuencia (RFID) consiste en un sistema de almacenamiento y recuperación de datos remoto que usa dispositivos denominados etiquetas (tags) (Figura 5). El propósito fundamental de esta tecnología es transmitir la identidad de un objeto mediante ondas de radio. Por tanto, se requiere que el vehículo lleve un dispositivo instalado (tag) capaz de comunicarse de forma segura y fiable con la infraestructura a través de un transmisor/receptor de comunicación (TRX).



Figura 5. Sensores de radiofrecuencia.

Por otro lado, un Sistema Global de Navegación por Satélite es una constelación de satélites que transmite rangos de señales utilizados para el posicionamiento y localización en cualquier parte de nuestro planeta (Figura 6). Estos sistemas permiten determinar las coordenadas geográficas y la altitud de un punto dado como resultado de la recepción de

señales provenientes de dichas constelaciones. Concretamente, este tipo de tecnología resulta muy útil para localizar con precisión y en todo momento un vehículo dado, conociendo el número de kilómetros recorridos, su velocidad media, punto de entrada y salida en la autopista, etc. La idea básica de todos estos sistemas es la triangulación. A partir de al menos tres satélites de posición conocida, y a partir de la distancia de cada uno de ellos al vehículo a localizar, es posible obtener su posición en términos de longitud y latitud.

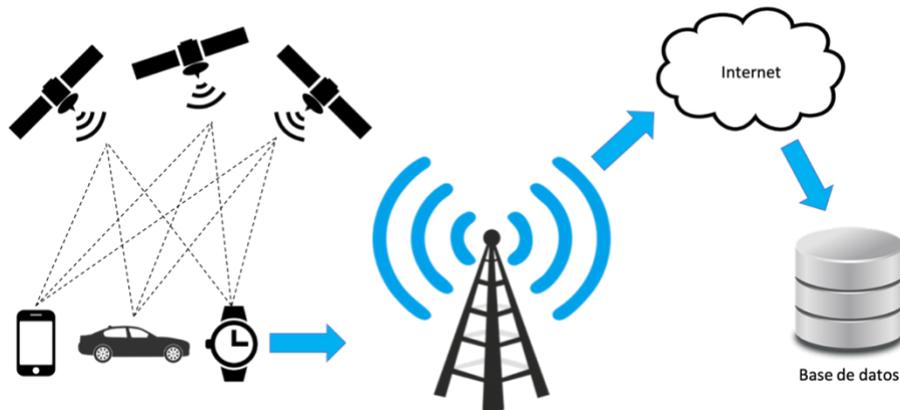


Figura 6. Sensores basados en tecnología GPS en entorno interurbano.

6 Sensores en entornos urbanos

En entornos urbanos es muy frecuente el uso de sensores pasivos para la detección y caracterización de la movilidad, esto es, sensores que no requieren que el usuario interactúe de una manera directa con el sensor para ser detectado por el mismo. Esto hace que los sensores recojan un gran volumen de datos de una manera objetiva a un coste muy reducido. No obstante, algunos de estos sensores requieren que los usuarios dispongan de ciertos dispositivos para ser detectados como, por ejemplo, sensores basados en bluetooth.

Entre los sensores más empleados cabe destacar el uso de redes inalámbricas, tecnología GPS, redes telefónicas y cámaras de vídeo (estudiados anteriormente). La gran cantidad de datos almacenados por estos sensores hace que la principal limitación de su uso esté en la reducción y el filtrado de los datos. Asimismo, el análisis de los datos también presenta dificultades debido a la ausencia de métodos estándar para ello.

Antes de pasar a estudiar cada uno de estos sensores, es interesante que conozcas los diferentes factores a considerar cuando trates con alguna de estas fuentes de datos:

- Tamaño de muestra: relación entre las personas que proporcionan información y el total de la población objeto de interés.
- Alcance espacial: ámbito de estudio, generalmente muy extenso.
- Precisión espacial: incertidumbre sobre la localización del dato. Puede ir desde varios metros hasta kilómetros
- Granulometría temporal: número de registros de un usuario por unidad de tiempo.
- Disponibilidad de datos sociodemográficos
- Otras limitaciones: necesidad de implantar infraestructura, software, batería...

6.1 Redes inalámbricas bluetooth

La tecnología inalámbrica bluetooth permite comunicaciones de corto alcance y de alta precisión enfocadas a dispositivos tanto móviles como fijos, manteniendo niveles altos de seguridad (Figura 7). Por tanto, su granularidad temporal es baja, puesto que solo se detectan mientras los dispositivos están en el área de influencia. Entre sus principales ventajas destacan su robustez, su baja potencia y su bajo coste. Si a esto le añadimos su elevado nivel de implantación y uso actual, se convierte en una tecnología muy interesante. En cuanto a los dispositivos personales más comunes detectados por este tipo de sensores, cabe destacar los teléfonos móviles.



Figura 7. Sensores basados en tecnología inalámbrica bluetooth.

Este tipo de sensores no indican la posición exacta del dispositivo detectado, sino que registran si un determinado dispositivo está dentro de su alcance. Además, la precisión de la caracterización de cierto movimiento dependerá fundamentalmente de la densidad de receptores inalámbricos dispuestos en la red (Figura 8). Aunque en la red representada en la Figura 8a podemos conocer que un determinado dispositivo se ha desplazado de A a B, no sabemos qué ruta escogió. Este problema podemos resolverlo aumentando el número de sensores (Figura 8b), en función del objetivo de nuestro estudio.

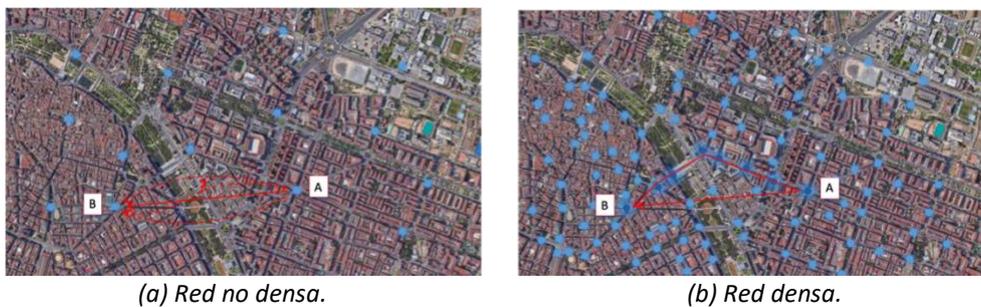
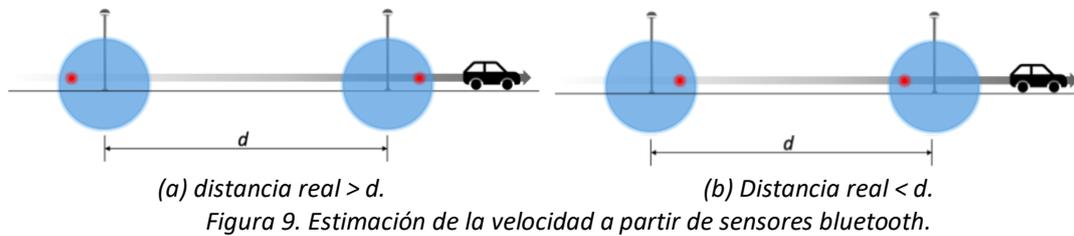


Figura 8. Densidad de la red basado en sensores bluetooth.

Es necesario remarcar que estos sensores requieren que el usuario tenga activo y visible el sistema bluetooth. Además, puede darse el caso de que un único usuario lleve consigo varios dispositivos conectados al mismo tiempo (móvil, reloj inteligente, tablet...), lo que podría sesgar el estudio. Por último, en caso de que queramos estimar la velocidad media de cierto usuario cuando hemos detectado su paso por dos sensores, debemos tener en cuenta la precisión de dicha estimación (Figura 9). Por ejemplo, si el vehículo ha recorrido una distancia mayor (la cual es desconocida) que la que realmente existe entre los sensores (d) (Figura 9a), estaremos subestimando su velocidad (calculada como el cociente entre d y la diferencia de tiempos de paso). En caso contrario (Figura 9b), la velocidad estimada será superior a la real. Para minimizar este error, debemos ser conscientes de que cuanto mayor sea la distancia entre los sensores, mayor será la precisión de la estimación.



Ya sabes cómo funcionan los sensores bluetooth y cómo almacenan sus datos. En el siguiente vídeo aprenderás a tratar estos datos y determinar, entre otras variables, la velocidad media con la que un usuario se ha desplazado entre dos puntos equipados con este tipo de sensor:

<https://media.upv.es/player/?id=031af850-69f3-11ea-ab9b-6df6378bebd6>

Si además quieres practicar, el archivo de datos lo puedes descargar haciendo click [aquí](#).

6.2 Redes de telefonía

Las redes de telefonía se caracterizan por:

- Gran tamaño muestral: el número de móviles supera al de habitantes.
- Alta cobertura de la red: es posible geoubicar a casi todos los usuarios.
- Conexión constante con la red: la conexión es continua gracias tanto a eventos activos (llamada, sms o sesión de datos) o pasivos (cambio de antena).
- Granularidad temporal muy alta: se obtienen infinidad de datos de un solo usuario.

A pesar de que las características de este tipo de sistema de recogida de datos nos lleven a pensar que puede tratarse de un sistema muy valioso, lo será realmente dependiendo del tipo de estudio que queramos llevar a cabo y la localización del ámbito de estudio. En este sentido, no se dispone de la localización exacta del dispositivo, sino qué dispositivos están conectados a una determinada antena. De este modo, la precisión de la geoubicación del dispositivo depende de la antena a la que está conectado y, generalmente, la densidad de estas antenas es mucho mayor en entornos urbanos que en rurales (Figura 10). De hecho, aunque los dispositivos móviles tienden a conectarse a la antena más cercana, en algunos casos pueden conectarse a otras debido a que la más cercana se encuentre muy saturada (elevado número de usuarios conectados).

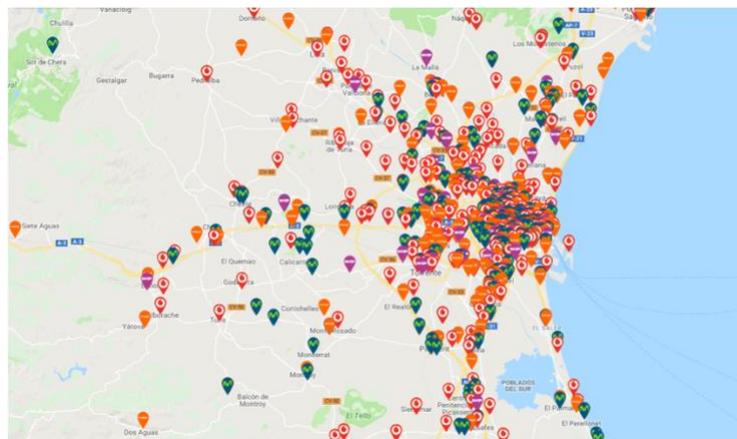


Figura 10. Red de antenas telefónica en la provincia de Valencia.

6.3 Sistemas basados en tecnología GPS

Los navegadores de conducción están basados en tecnología GPS. Como ya se ha comentado anteriormente, este tipo de tecnología resulta muy útil para localizar con precisión y en todo momento un vehículo dado y su velocidad media (Figura 6). Además de estos navegadores, también los teléfonos móviles pueden hacer uso de esta tecnología a través de aplicaciones instaladas y activadas a voluntad del usuario. En este sentido, muchas de ellas solo necesitan ser activadas por primera vez y posteriormente se ejecutan en segundo plano. Un ejemplo de ello son las aplicaciones de Google, que utilizan nuestros datos de geolocalización para estimar el nivel de tráfico en un determinado momento e, incluso, para realizar predicciones relacionadas con la hora a la que llegaremos a un determinado lugar.

Un caso particular del uso de la tecnología GPS son las redes sociales, principalmente, destinadas al registro de actividades deportivas como Strava, Endomondo o Runtastic. Estas redes sociales requieren que el usuario se registre, por lo que habitualmente recogen variables sociodemográficas que podrían ser de gran interés para nosotros. Una vez registrado, el usuario puede cargar de manera voluntaria todas aquellas actividades deportivas que desee. Aunque su representatividad en entorno urbano es reducida, en entornos interurbanos puede llegar a dar muy buenos resultados. Finalmente, cabe destacar que la precisión de los datos disponibles depende fundamentalmente del dispositivo y/o la aplicación utilizada y, algunas de las aplicaciones, permite recoger datos relacionadas con el pulso cardiaco o la cadencia del paso.



Observa el mapa de calor de la red social Strava para actividades realizadas a pie (senderismo y carrera). ¿Observas algo especial? ¡Sí, se pueden identificar perfectamente las diferentes rutas del Camino de Santiago! Si quieres explorar más sobre este tipo de mapas, visita:

<https://www.strava.com/heatmap>



7 Conclusión

En este documento hemos estudiado los distintos sistemas de recogida de datos masivos en relación con la ingeniería de tráfico. En cuanto a entorno interurbano, los sistemas se pueden clasificar en autónomos, ya sean intrusivos o no intrusivos, y dependientes. Por otro lado, los sistemas más empleados en entorno urbano son las redes inalámbricas, principalmente basadas en tecnología bluetooth, las redes telefónicas y los sistemas basados en tecnología GPS. Asimismo, las encuestas pueden ser empleadas en ambos tipos de entornos para conseguir, además de información objetiva, también las opiniones de los usuarios de la red.

8 Bibliografía

Plataforma Tecnológica de la Carretera (2011). "Sistemas de adquisición de información de tráfico: Estado actual y futuro", 01/2011, Plataforma Tecnológica de la Carretera, Madrid.