

## Detección de modos de transporte usando datos GPS<sup>δ</sup>

### *Transportation Mode Detection using GPS Data*

Marco Antonio Adarme Jaimes\*  
Diana Heredia Vizcanio\*\*  
Eduard Gilberto Puerto Cuadros\*\*\*



Recibido: 21 de noviembre de 2016

Aceptado: 27 de enero de 2017

Tipo de artículo: resultado de investigación

### Resumen

El uso de dispositivos móviles y el aprovechamiento de la tecnología GPS, permiten la implementación de sistemas para analizar el contexto y actividades típicas de transporte de un usuario, a través del análisis de los datos de localización y sensores de aceleración. Este trabajo de investigación comprende el procesamiento de datos obtenidos vía GPS. Con este procesamiento se pretende detectar el modo de transporte de un usuario en segmentos de recorridos predefinidos. Para la clasificación de éstos, se usan perfiles de velocidad que identifican los modos de transporte en cada uno de los segmentos, mediante un sistema software en lenguaje de programación Java y la utilización de Matlab para el análisis y filtros de datos. El sistema software se desarrolla en dos componentes, el primero comprende el filtro y transformación de datos. Estos datos se representan en coordenadas decimales a coordenadas cartesianas. El segundo presenta la clasificación, para la detección de modos de transportes con las coordenadas cartesianas. También contiene el análisis de estados de movimientos cinemáticos. Las pruebas se realizan a través de un *dataset* tomado del proyecto *GeoLife* de *Microsoft Asia*. Los resultados obtenidos muestran una detección coherente sobre los medios de transporte que usan los diferentes usuarios. Estos usuarios se comparan a partir de perfiles de velocidad predefinidos.

**Palabras Clave:** Detección de modos de transporte, GPS, transporte multimodal.

### Abstract

The use of mobile devices and GPS technology allow the implementation of systems to analyze the context and typical transport activities of a user, through the analysis of the location data and acceleration sensors. This research includes the processing of data obtained via GPS. This processing is intended to detect the mode of transport of a user in segments of predefined paths. For classification, velocity profiles that identify modes of transport in each segment are used. The software implements a Java programming language and the use of Matlab for analysis and data filters. The software system is developed into two components; the first comprises the filter and transformation of data. These data are plotted from decimal coordinates to cartesian coordinates. The second presents the classification for the detection of transport modes with cartesian coordinates. It also contains the analysis of states of kinematic movements. The tests are performed through a dataset taken from the GeoLife project of Microsoft Asia. The obtained results show a coherent detection on the means of transport that the different users use. These users are compared from predefined speed profiles.

**Keywords:** Transportation mode detection, GPS, multimodal transport.

Este artículo es resultado del proyecto de investigación Clasificación de medio de transporte a partir de secuencias de seguimiento utilizando datos GPS, del Grupo de Investigación y Desarrollo de Ingeniería de Software de la Universidad Francisco de Paula Santander (Cúcuta-Colombia) y el grupo de Investigación IngeBioCaribe de la Universidad Simón Bolívar(Barranquilla-Colombia).

\* MSc. Profesor Departamento de Sistemas e Informática-Universidad Francisco de Paula Santander, Cúcuta-Colombia. Correo Electrónico: madarme@ufps.edu.co

\*\* MSc. Profesor Facultad de Ingeniería. Universidad Simón Bolívar, Barranquilla-Colombia. Correo Electrónico: dianahv@unimonbolivar.edu.co.

\*\*\* MSc. Profesor Departamento de Sistemas e Informática-Universidad Francisco de Paula Santander, Cúcuta-Colombia. Correo Electrónico: eduardpuerto@ufps.edu.co

## Introducción

El uso de dispositivos móviles inteligentes se ha incrementado de manera exponencial, ellos están dotados de hardware altamente especializado con funciones de captura datos tales como aceleración, localización, imágenes, entre otros. Estos datos se pueden usar de manera automática o a partir de alguna interacción (E-Interactive, 2013). A través de este hardware especializado, los dispositivos móviles pueden convertirse en grandes sistemas de sensores de propósito general o específico permitiendo la toma de datos a lo largo de un período y abarcando grandes zonas geográficas con el uso de la red celular. En términos generales, el sistema se modela como una red de sensores altamente distribuidos que trabajan en la adquisición de datos en una zona geográfica en particular.

Con estos aspectos y apoyados en áreas de investigación como computación ubicua y computación consciente del contexto, los dispositivos móviles están siendo usados para obtener información de su ambiente como indicadores de movilidad y preferencias en uso de transportes, que resulta de vital importancia para sistemas de información urbanos que necesitan de toda estos datos para sus procesos de control y estudio de una determinada ciudad (Stenneth, Wolfson, Yu, & Xu, 2011).

Este trabajo comprende el estudio de los datos obtenidos del GPS de dispositivos móviles y la forma de determinar y procesar variables que permitan detectar el modo de transporte en una actividad demarcada por inicio y paradas del usuario en un tiempo determinado (movimiento cinemático), para esto se plantean tres objetivos:

Desarrollar un sistema software para el procesamiento y filtro de datos GPS para la identificación de los modos de transporte de un usuario.

Implementar una técnica para la detección de modos de transporte, a través de perfiles de velocidad predefinidos, valores de movimiento cinemático y de localización.

Realizar pruebas sobre un banco de datos de propósito general.

Para el desarrollo de los dos primeros objetivos se utilizaron estudios que demostraban la efectividad de utilizar perfiles de velocidad, tiempo y localización para la predicción de modos de transporte; para el último objetivo se toma un *dataset* del proyecto *GeoLife de Microsoft Asia* (Microsoft Research, 2012) que obtienen datos GPS de 178 usuarios en intervalos de 1 a 5 segundos.

El artículo se organiza de la siguiente manera: en la sección 2 se presenta la revisión de literatura, donde se abordan las investigaciones y proyectos relacionados con detección de modos de transporte, análisis de datos GPS y manejo de sensores urbanos, entre otros conceptos contenidos en el área de transporte multimodal. En la sección 3 se define la metodología de trabajo para el filtro y conversión de coordenadas decimales a coordenadas cartesianas y la interacción del sistema software para la detección de los modos de transporte. Finalmente se exponen los resultados obtenidos del filtro y de procesamiento de coordenadas cartesianas, así como de la detección de modos de transporte.

## Revisión de literatura

La investigación en curso, toma como referencia estudios sobre la predicción de los modos de transporte a través del procesamiento de datos relevantes que alimenten un modelo de clasificación, así como estudios concernientes a la captura, filtraje y procesamiento de datos GPS.

En (Shin, Aliaga, Tunçer, Müller, Kim, & Zünd, 2014) se crea un prototipo de aplicación sobre plataformas Android que toma datos del sensor de aceleración, de localización y permite clasificar los medios de transporte. El proceso consiste en identificar los segmentos de un transporte en particular basado en perfiles de velocidad (información que se obtiene con la aceleración y la localización), los segmentos se identifican a través de puntos adyacentes que comprenden el inicio y fin de un recorrido, estos puntos se toman de la acción de cambio de velocidad y de la predicción

de la actividad de caminar de un usuario (velocidad menor o igual a 7.2km/h).

De igual forma los trabajos propuestos por (Zheng, Liu, & Wang, 2008) y (Zheng, Chen, Quanan, Xie, & Wei-Ying, 2010) analiza las variables GPS que pueden ser usadas en la identificación de los modos de transporte y utiliza algoritmos de aprendizaje dinámico que van alimentando el sistema para la inferencia automática de estos. El proceso consiste en dividir las actividades de un usuario en segmentos donde se identifican cambios considerables de posición y velocidad. Seguido de esto, cada segmento se dota de características de su contexto usando otros sistemas de información geográfica. Estos sistemas permiten tener predicciones más reales y sirven como entrada para un modelo inferencial que clasifica los modos de transporte. Finalmente, el sistema aprende del comportamiento de un usuario con el uso prolongado de la aplicación y detecta comportamientos típicos en rutas y movimientos con mayor exactitud.

En (Zhan, Hasan, Ukkusuri, & Kamga, 2013) se determinan rutas, tiempo, y tráfico de un sistema de taxis basados en datos obtenidos vía GPS, comportándose como una gran red de sensores. El modelo toma cada ruta de taxi, identificando puntos de paradas preestablecidos y su cercanía, y calcula las posibles rutas donde el taxi pasará, los conjuntos de camino generados sirven como base para el proceso de estimación de su tiempo de viaje, se predice la velocidad de llegada entre rutas a través de un modelo matemático basado en grafos. Se diseña un sistema centralizado de adquisición de datos y se demuestra la efectividad de usar dispositivos móviles en vez de los sensores tradicionales.

En Oriol *et al.* (2013) se propone un sistema para analizar movilidad turística en la ciudad de Taragona, España usando datos recolectados vía GPS. Para la recolección de datos se cuenta con equipos GPS que son entregados a los turistas y estos son devueltos al cabo de un tiempo, de igual forma se analizan los perfiles de los visitantes a través de entrevista previas.

El proyecto *Semantic trajectories: Mobility data computation and annotation* (Yan, Chakraborty, Parent, & Spaccaprieta, 2013), los autores desarrollaron una plataforma para agregar valor semántico a las trayectorias identificadas a partir de datos de localización capturados por medio de GPS desde diferentes tipos de sensores y dispositivos móviles. Una trayectoria semántica es una forma de enriquecer los datos primarios, añadiendo información del contexto. La plataforma consta de tres capas principalmente: El modelo de datos primarios: Que corresponde a los registros espacio-temporales que provienen directamente de los dispositivos dotados con sistemas GPS, el modelo conceptual: divide de forma lógica una trayectoria espacio-temporal en una serie de episodios no superpuestos y el modelo Semántico: Las trayectorias estructuradas se enriquecen con anotaciones semánticas, gracias a la integración de datos externos, tales como bases de datos geográficas o datos de localización en redes sociales.

En (Ashbrook & Starner, 2003) se presenta una metodología para la recolección de datos y un modelo para predecir rutas a través de GPS, su trabajo se centra en convertir estos datos a través de un sistema de Markov (modelo estadístico donde el sistema toma como entradas datos probabilísticos y genera posibles datos basados en reglas de inferenciales predefinidas). Su sistema toma dos estrategias para la recolección de la localización, la primera utiliza un muestreo de usuarios de manera individual y la otra toma los datos de todos los usuarios como una gran red de sensores.

Otros sistemas incorporan la utilización de computación inteligente para analizar datos GPS, tal es el caso del proyecto *Adaptive personalized travel information systems* (Arentze, 2013) donde se utiliza un método bayesiano para la predicción de rutas y preferencias de transporte de viajeros con algoritmos que tienen la facultad de aprender usando una base de conocimiento dinámica alimentada por el comportamiento de los usuarios en un determinado periodo de tiempo.

En (Zhang, Arentze, & Timmermans, 2012), se desarrolla un sistema de información para viajeros basado en un modelo genérico de identificación de transporte multimodal. Su modelo toma como entrada un grafo dirigido donde cada nodo es una posición física y sus enlaces representan eventos que almacenan el tiempo de un punto a otro según el tipo de transporte que se valla usar. El sistema proporciona una base de datos para que las aplicaciones móviles a través de datos GPS puedan operar sobre su modelo.

El proyecto realizado por Xio *et al.* (2015) implementa un sistema sobre una red bayesiana sobre algoritmos de aprendizaje K2 que permiten extraer información de conocimiento previo y utilizan técnicas avanzadas de minería de datos. El sistema extrae la información de datos GPS y a través de técnicas de similitud se estiman los modos de transporte. La red identifica con gran precisión los cambios de tasas de velocidad y de trayectorias en rutas, siendo capaz de ajustar los parámetros GPS en factores de similitud para la detección adecuada del modo de transporte.

En (Wu, Yang, & Jing, 2016) realiza una revisión sobre los métodos formales para la recolección de datos GPS y detección de modos de transporte proponiendo una metodología de adquisición, filtro y procesamiento de datos a través de plataformas android y los posibles algoritmos de detección de modos de transporte. Su propuesta ofrece un panorama del sistema desde la perspectiva del cliente y otra del servidor. El sistema se basa en procesos en tiempo real y el como se podrían optimizar los algoritmos y el consumo energético en los smartphones. Así mismo Feng *et al.* (2016) presenta una comparación sobre los algoritmos utilizados para extraer patrones de aceleración que identifican modos de transporte sobre segmentos con límites de longitud y tiempo. Su estudio muestra una taxonomía de algoritmos basados en redes bayesianas, regresiones lineales, *vector machine* y tablas de decisión. La investigación da como resultado que los

algoritmos sobre redes bayesiana tienen un mejor desempeño para la detección de modos de transporte debido a que pueden procesar y filtrar una gran cantidad de datos en un tiempo de complejidad menor en comparación de los otros algoritmos.

En (Zhu, Li, Liu, Wang, & Yang, 2016) desarrollan un sistema de inferencia de modos de transporte con algoritmos en redes neuronales basados en datos GPS del proyecto Geolife de Microsoft Asia (Microsoft Research, 2012). Su método consiste en personalizar los perfiles de velocidad de cada modo de transporte de tal manera que el mismo sistema aprenda si existe variación de velocidad y pueda clasificar cada actividad de recorrido de acorde al mejor medio de transporte que se ajuste. El aporte principal de esta investigación es un *framework* para la construcción de perfiles de velocidad de movilidad de usuarios con datos dispersos y sin filtrar GPS basados en restricciones de las características de velocidad de una red de transporte.

## Metodología

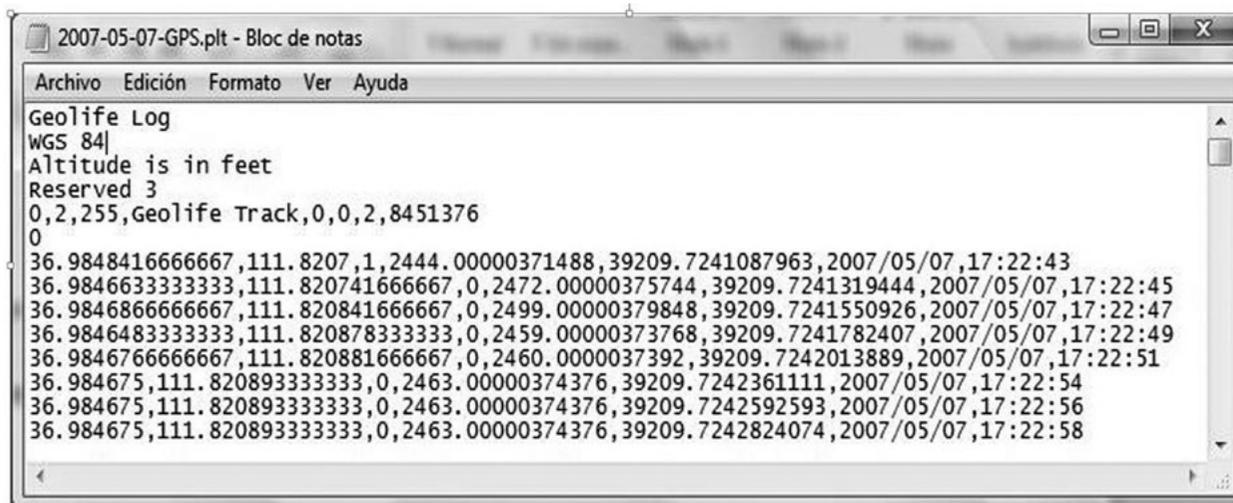
Para el desarrollo del análisis de datos y clasificación de los modos de transporte usados por los usuarios, se realizaron las siguientes actividades:

### 1. Obtención de datos primarios:

Para la realización de este trabajo se utilizó un GPS trajectories dataset, obtenido a través del proyecto Geolife de Microsoft Asia (Microsoft Research, 2012); consta de datos de localización de 178 usuarios, recolectados entre abril de 2007 y octubre de 2011. Los datos en cada archivo corresponden a una sucesión de puntos capturados mediante diversos dispositivos dotados de GPS, con intervalos de tiempo entre 1 y 5 segundos.

Los campos en cada archivo son:

- Longitud y latitud (en grados decimales)
- Altitud (en pies)
- Timestamp (convertido además en fecha y hora en campos separados).
- La figura 1 muestra un ejemplo de estos datos.



**Figura 1.** Ejemplo de Datos primarios.

Fuente: Adaptación Microsoft Research (2012)

Los medios de transporte reportados, aunque este dato no aparece en los archivos, son:

- Automóvil (taxi o particular)
- Bus
- Bicicleta
- Tren ligero (ligh train, no subterráneo)

## 2. Procesamiento de los datos:

Tal como se muestra en la figura 1, los datos primarios se encuentran en un formato plt, separado por comas y se prepararon en Excel, de la siguiente forma:

- a. Conversión de coordenadas geográficas a coordenadas cartesianas: Es necesario obtener las coordenadas de cada punto, las cuales corresponden a la distancia en metros desde el meridiano cero para la longitud y desde la línea del Ecuador para la latitud, respectivamente. Esta conversión se realizó utilizando un programa en Matlab, *deg2utm.m*, obtenido a través de Mathworks (Palacios, Rafael, 2006), el cuál implementa las fórmulas sobre el elipsoide de Hayford (Featherstone & Claessens, 2008). Los datos de longitud y latitud se cargaron a Matlab desde Excel como vectores separados y se pasaron al programa *deg2utm* y grabados los resultados por éste en otro archivo en Excel.

- b. Cálculo de diferencia de parámetros de trayectoria: Con las coordenadas cartesianas, se procede a calcular los siguientes parámetros, también en Excel:

- Diferencia de tiempo: Se calculó la diferencia entre dos timestamp consecutivos y convertido a su vez a segundos, utilizando las funciones Minutos y Segundo de Excel.
- Distancia recorrida: Se calculó entre dos puntos consecutivos, por medio de la fórmula de distancia euclidiana y la distancia se da metros.
- Velocidad instantánea: Corresponde al valor de distancia recorrida entre dos puntos sucesivos entre la diferencia de tiempo entre ellos ( $v=d/\Delta t$ )
- Aceleración instantánea: Corresponde al valor de diferencia de velocidades entre dos puntos sucesivos entre la diferencia de tiempo ( $A=\Delta v/\Delta t$ ).

- c. Identificación de segmentos: Un segmento es un intervalo de tiempo donde el usuario está detenido (STOP), en el cuál la velocidad es cero en varios puntos consecutivos, o en movimiento (NO STOP), si la velocidad es diferente de cero. Para cada punto de la trayectoria se marcó si estaba detenido o no, por medio

de una función condicional en Excel. Una muestra de los datos obtenidos de este proceso, se puede observar en la figura 2. Posteriormente, los datos completos son procesados a través de una aplicación en Java, que los ubica en un modelo de clases (ver figura 3) de trayectorias y segmentos de la siguiente forma:

- Varios puntos consecutivos en estado detenido (o no detenido) forman un segmento.
- Para cada uno de estos segmentos se calcularon los siguientes parámetros: Distancia total recorrida, velocidad promedio y aceleración promedio.

LATITUD	LONGITUD	CORRECTO	ALTITUD	TIMESTAMP	FECHA	HORA	DISTANCIA	DIF TIEMPO	VELOCIDAD	ACELERACION	STOP
573037,848	4093505,5	1	2444	39209,7241	07/05/2007	05:22:43 p.m.					
573041,727	4093485,75	0	2472	39209,7241	07/05/2007	05:22:45 p.m.	20,1288845	2,000000000	10,0644422		NO STOP
573050,604	4093488,42	0	2499	39209,7242	07/05/2007	05:22:47 p.m.	9,26879044	2,000000000	4,63439522	-2,71502351	NO STOP
573053,904	4093484,19	0	2459	39209,7242	07/05/2007	05:22:49 p.m.	5,36048823	2,000000000	2,68024411	0,97707555	NO STOP
573054,173	4093487,34	0	2460	39209,7242	07/05/2007	05:22:51 p.m.	3,15727604	2,000000000	1,57863802	-0,55080305	NO STOP
573055,213	4093487,16	0	2463	39209,7242	07/05/2007	05:22:54 p.m.	1,05466519	3,000000000	0,35155506	0,40902765	NO STOP
573055,213	4093487,16	0	2463	39209,7243	07/05/2007	05:22:56 p.m.	0	2,000000000	0	-0,17577753	STOP
573055,213	4093487,16	0	2463	39209,7243	07/05/2007	05:22:58 p.m.	0	2,000000000	0	0	STOP
573055,213	4093487,16	0	2463	39209,7243	07/05/2007	05:23:00 p.m.	0	2,000000000	0	0	STOP
573055,213	4093487,16	0	2463	39209,7243	07/05/2007	05:23:02 p.m.	0	2,000000000	0	0	STOP
573055,213	4093487,16	0	2463	39209,7244	07/05/2007	05:23:04 p.m.	0	2,000000000	0	0	STOP
573055,213	4093487,16	0	2463	39209,7244	07/05/2007	05:23:06 p.m.	0	2,000000000	0	0	STOP
573064,897	4093482,25	0	2452	39209,7244	07/05/2007	05:23:08 p.m.	10,8574628	2,000000000	5,4287314	2,7143657	NO STOP
573066,4	4093480,05	0	2453	39209,7244	07/05/2007	05:23:10 p.m.	2,66896353	2,000000000	1,33448177	-2,04712482	NO STOP
573068,052	4093477,66	0	2450	39209,7244	07/05/2007	05:23:12 p.m.	2,90519004	2,000000000	1,45259502	0,05905663	NO STOP
573069,242	4093477,3	0	2451	39209,7245	07/05/2007	05:23:14 p.m.	1,24295075	2,000000000	0,62147537	-0,41555982	NO STOP
573069,987	4093476,94	0	2451	39209,7245	07/05/2007	05:23:16 p.m.	0,82874643	2,000000000	0,41437321	-0,10355108	NO STOP

Figura 2. Muestra de datos preprocesados.

Fuente: Elaboración Propia (2016)

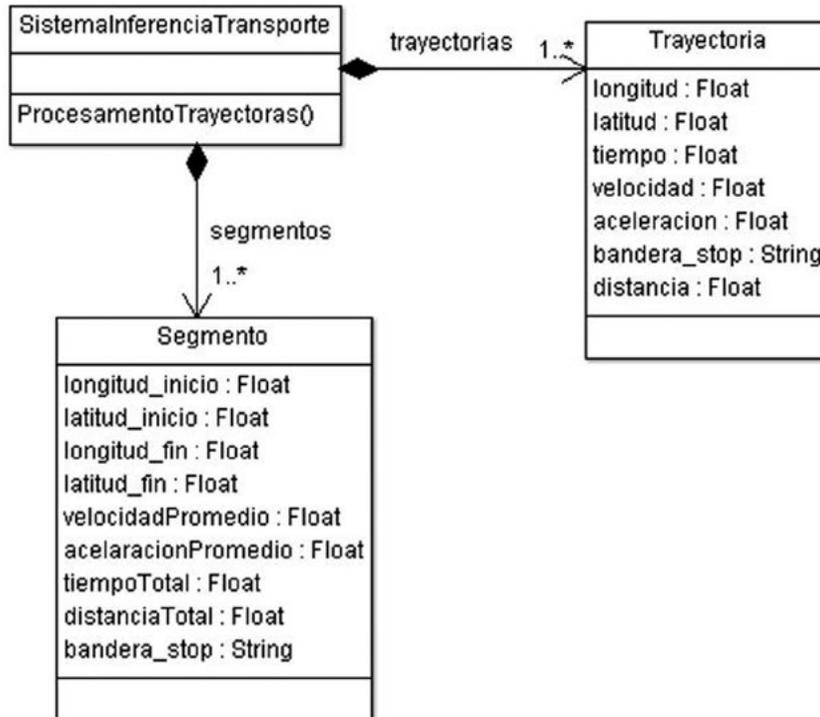


Figura 3. Modelo de clases de trayectorias y segmentos.

Fuente: Elaboración Propia (2016)

- Además, se almacenaron las coordenadas del punto de inicio del segmento y las del punto final.
- d. Estrategia de análisis para la detección de segmentos: Como se menciona en el punto anterior cada segmento se toma desde una trayectoria inicial leída del archivo plano que contiene los datos preprocesados con las coordenadas cartesianas y valores cinemáticos. Debido a la gran cantidad de datos que puede contener los diferentes *dataset* se implementa un modulo de filtro de datos que permite entre otros aspectos, depurar las trayectorias para crear solo los

segmentos que tienen información relevante que ofrezca la detección del modo de transporte por parámetros de similitud de perfiles de velocidad presentados en la tabla 1. Es común encontrar datos que contienen muchas banderas de “STOP” para esto se realiza el cálculo del tiempo de inercia y se promedia de tal manera de solo tener una bandera de “STOP” y otra de “NO STOP” por cada segmento que se identifique. El segmento se crea con la identificación de la longitud recorrida con valores de velocidad y aceleración y su variación es comparada y ajustada con los intervalos de velocidad establecidos en la tabla 1.

**Tabla 1.** Velocidades típicas de diversos modos de transporte

Modo de transporte	Velocidad promedio (Km/h)	Velocidad promedio (m/s)
Caminar	3 a 4,5	0,8 a 1,5
Bicicleta	6 a 20	1,5 a 5,5
Automóvil	30 a 50	8,3 a 14
Bus urbano	20 a 30	5,5 a 8,3

Fuente: Adaptación de Ecomovilidad (2012).

De esta forma, la estructura de datos que almacena los segmentos solo contiene datos relevantes para detectar el medio de transporte que más se ajuste a un segmento en particular.

Cabe resaltar la importancia de la etapa de depuración de trayectorias y de identificación de segmentos, ya que el archivo con las trayectorias iniciales contiene un sin número de valores que no son relevantes para la creación de segmentos y que son propios de la toma de datos que ofrece el GPS y de la fluctuación de tiempo en las lecturas, esta idea de procesamiento y depuración de datos se estudia en (Zheng, Chen, Quanan, Xie, & Wei-Ying, 2010) y en (Xiao, Juan, & Zhang, 2015).

Gracias a esto, el archivo plano inicial del *dataset* que pesaba una cantidad de MBytes considerable, se pudo reducir y colocar sobre una estructura de datos en forma dinámica que permitió optimizar el procesamiento del sistema.

- e. Estimación del modo de transporte: Utilizando el modelo de clases construido por la aplicación en Java, se determina el modo de transporte para cada segmento a partir de la tabla de perfiles de velocidad basados en valores mínimo y máximo típicos en la movilidad de una ciudad (ver tabla 1) basados en (Ecomovilidad, 2012). Para efectos de normalizar las unidades, los rangos de velocidad se tomaron en metros por segundo. Los resultados del proceso de segmentación e identificación de medios de transporte se pueden observar en la figura 3.

### Resultados

Se tomó una muestra de los datos consignados en el dataset descrito, con 2360 registros, los cuales se sometieron al procesamiento antes descrito y se obtuvieron los resultados mostrados a continuación (ver figura 4).

Segmento	Latitud inicial	Latitud Final	Longitud Inicial	Longitud Final	Distancia Total	Dif Tiempo Total	Velocidad Promedio	Aceleración promedio	Tipo Transporte
NO STOP	573041,75	573041,75	4093485,8	4093487,2	38,970104	11	.861855	-0,93038595	Bicicleta
STOP	573055,2	573055,2	4093487,2	4093487,2	0	12	0	-0,029291255	Ninguno
NO STOP	573064,9	573064,9	4093482,2	4093477	18,503313	10	16503313	0,041437317	Bicicleta
STOP	573070	573070	4093477	4093477	0	4	0	-0,103593305	Ninguno
NO STOP	573074	573074	4093475,8	4093472,8	9,006823	9	1,0145267	0,08413952	Caminando
STOP	573077,44	573077,44	4093472,8	0	0	2	0	-0,13867533	Ninguno
NO STOP	573076,25	573076,25	4093476	4093427,2	122,25411	106	1,1500583	0,006512588	Caminando
STOP	573148,44	573148,44	4093427,2	4093427,2	0	6	0	-0,06906254	Ninguno
NO STOP	573147,3	573147,3	4093434,8	4093436,5	9,665594	4	2,4163985	0,2746306	Bicicleta
STOP	573148,5	573148,5	4093436,5	4093436,5	0	14	0	-0,07846589	Ninguno
NO STOP	573145,1	573145,1	4093433,5	4093434	6,7884417	9	0,8687247	0,15034446	Caminando
STOP	573147,2	573147,2	4093434	4093434	0	10	0	-0,1074741	Ninguno
NO STOP	573150,94	573150,94	4093433,2	4093435,2	6,00149	4	1,5003725	0,27751338	Bicicleta
STOP	573151,8	573151,8	4093435,2	4093435,2	0	13	0	-0,092504464	Ninguno
NO STOP	573153,8	573153,8	4093440,8	4093440,8	6,219,668	4	1,554,917	0,037083447	Bicicleta
STOP	573154,1	573154,1	4093440,8	4093440,8	0	4	3	-0,037083447	Ninguno
NO STOP	573143,56	573143,56	4093427,5	4093443	79,85662	84	4,08779	0,41128102	Bicicleta
STOP	573157,06	573157,06	4093443	4093443	0	5	0	-0,4165368	Ninguno
NO STOP	573155,7	573155,7	4093445,8	4093439,8	21,052881	19	1,1835212	0,06800694	Ninguno
STOP	573149,8	573149,8	4093439,8	4093439,8	0	8	0	-0,0929139	Ninguno
NO STOP	573149	573149	4993429,2	4093425	14,8249655	8	1,375523 2	-0,2374035	Bicicleta
STOP	573148,9	573148,9	4093425	4093425	0	9	0	-0,05852041	Ninguno
NO STOP	573148,3	573148,3	4093407,2	4093424,8	68,12512	404	3,7393932	-0,09807998	Caminando
STOP	573153,2	573153,2	4096124,8	4093424,8	0	6	0	-0,11865183	Caminando
NO STOP	573146	573146	4093417,8	4093442,2	63,79288	52	12,319,179	0,016877428	Caminando
STOP	573112,7	573112,7	4093442,2	0	0	2	0	-0,19109239	Ninguno
NO STOP	573109,9	573109,9	4093444,2	4093457,8	41,011032	37	1,099E+139	-0,001635287	Ninguno
STOP	573084,56	573084,56	4093457,8	4093457,8	0	15	0	-0,017210195	Ninguno
NO STOP	573072,06	573072,06	4093461,5	4093483,2	38,437446	23	1,720446 8	0,039141856	Bicicleta
STOP	573060,44	573060,44	4093483,2	4093483,2	0	12	0	-0,048551336	Caminando
NO STOP	573051,2	573051,2	4093487,2	4093485,2	12,784986	6	7,130831	0,095698655	Bicicleta
STOP	573050,3	573050,3	4093485,2	0	0	2	0	-0,28709587	Ninguno
NO STOP	573044,1	573044,1	4093483	4093476,2	7,467,433	554	10,565978	2,1135287	Auto-Taxis
STOP	573041,4	573041,4	4093476,2	4093476,2	0	4	0	-0,73755551	Ninguno
NO STOP	573050,3	573050,3	4093491	4424257	1045053,75	18177	22,618095	-0,075274356	Otro medio
STOP	444009,88	444009,88	4424257	0	0	2	0	-1,6709815	Ninguno
NO STOP	444014,22	444014,22	4424268,5	4424336,5	34,075,613	57	6,804,353	1,0497918	Bus
STOP	444048,16	444048,16	4424336,5	0	0	2	0	-21,82277	Ninguno
NO STOP	444923,94	444923,94	4425705,5	4425737,5	1836,3413	982	70014358	-0,040668573	Bus
STOP	444857,44	444857,44	4425737,5	0	0	2	0	-0,86212466	Ninguno
NO STOP	444860,28	444860,28	4425740,5	4425742,5	12,646098	10	1,264609 8	0,089783326	Caminando
STOP	444859,47	444859,47	4425742,5	4425742,5	0	6	0	-0,14963886	Ninguno
NO STOP	444852,66	444852,66	4425747,5	4425748,5	11,529637	7	1,4399625	0,04310574	Caminando
STOP	444854,94	444854,94	4425748,5	4425748,5	0	8	0	-0,15296525	Ninguno
NO STOP	444844,6	444844,6	4425753,5	4425773,5	481,875	69	7,04675 5	0,0280147	Bus
STOP	445275,38	445275,38	4425773,5	4425773,5	0	6	0	-0,061267857	Ninguno
NO STOP	445273,12	445273,12	4425775	4425203,5	800,45166	150	7,5039425	0,21288879	Bus

Figura 4. Ejemplo de segmentos y modo de transporte identificados.

Fuente: Elaboración Propia (2016).

Como se puede observar, con la velocidad promedio de un segmento se puede determinar con algún grado de certeza, el modo de transporte utilizado por una persona. Sin embargo, la identificación de modo de transporte se basa en rangos de velocidad que fueron tomados sin superposiciones para simplificar al problema el análisis; en realidad, varios modos de transporte pueden coincidir en velocidad (por ejemplo, automóvil y bus).

Los segmentos con estado STOP pueden representar situaciones como detenerse en un semáforo o algún otro tipo de señal de tránsito, una parada en la ruta del bus, una pausa o permanecer en un sitio distinto de un medio de transporte durante un tiempo. Aparentes cambios de medio de transporte en medio de segmentos consecutivos y cuya duración es corta, no deberían ser considerados como reales. Por ejemplo, se identifican varios segmentos con modo de transporte “caminando” alternados con segmentos de no movimiento durante un tiempo relativamente largo, pero en medio de ellos hay un segmento con modo de transporte “bicicleta” con duración de pocos segundos; éste último segmento podría interpretarse como “caminando” también.

La secuencia de modos de transporte obtenida como resultado final de todo el proceso es bastante coherente en cuanto hay segmentos consecutivos en un medio de transporte dado, con pocos cambios abruptos aparentes.

## Conclusiones

A partir de los datos primarios de movilidad de usuarios, capturados desde sus dispositivos móviles, se puede obtener información diversa y útil acerca de los medios de transporte utilizados, el contexto ambiente donde fueron adquiridos, su comportamiento habitual, los sitios frecuentados, el momento del día, entre otros.

Fue necesario para el análisis y procesamiento de los datos GPS obtenidos del *dataset* del proyecto *GeoLife* de Microsoft Asia, etapas de depuración y preprocesamiento de datos para

encontrar los segmentos que ofrecen la identificación de movimiento con velocidad y aceleración promedio. Esta información se compara con las rangos de velocidad comunes a cada uno de los medios de transporte presentados en la tabla 1. La utilización de componentes *Matlab* para la conversión de coordenadas, agilizaron la etapa de preprocesamiento de datos y ofrecieron datos en coordenadas cartesianas que son de fácil interpretación.

La depuración de trayectorias iniciales del *dataset* y filtros que generaban solo segmentos con identificaciones de perfiles de movimiento y velocidad bien definidos, permiten obtener detecciones coherentes e información relevante que puede ser utilizada para toma de decisiones o como base para análisis de datos más complejos con respecto a la movilidad urbana en cualquier lugar.

El sistema se desarrolla con características configurables lo cual ofrece la ventaja de cambiar los valores en los estados de movimiento cinemático y de los perfiles de velocidad.

Como trabajos futuros se espera realizar una caracterización más amplia de los perfiles de velocidad y la automatización de procesos de lectura de datos GPS en una aplicación cliente que permita la lectura online de datos y su posterior procesamiento.

## Referencias

- Arentze, T. (2013). Adaptive personalized travel informationsystems: A bayesian method to learn users' personal. *IEEE Intelligent Transportation Systems Society*, 14(4), 1957 - 1966.
- Ashbrook, D., & Starner, T. (2003). Using GPS to learn significant locations and predict movement across. *Personal and ubiquitous computing*, 7(5), 275-286.
- Ecomovilidad. (2012). ¿A qué velocidad va el transporte público?. Recuperado de <http://>

- ecomovilidad.net/madrid/a-que-velocidad-va-el-transporte-publico/
- E-Interactive. (2013). Comportamiento de usuarios en dispositivos móviles. Recuperado de <http://www.e-interactive.es/wp-content/uploads/Informe-de-Marketing-Movil.pdf>
- Featherstone, W., & Claessens, S. (2008). Closed-form transformation between geodetic and ellipsoidal coordinates. *Springer*, 52(1), 1-18.
- Feng, T., & Timmermans, H. J. (2016). Comparison of advanced imputation algorithms for detection of transportation mode and activity episode using GPS data. *Transportation Planning and Technology*, 39(2), 180-194.
- Microsoft Research. (2012). *GeoLife GPS trajectories*. Recuperado de <http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/>
- Oriol , B., Bigorra, A., Pérez , Y., Orellana, A., Russo, A., & Salvador Anton, C. (2013). Analysis of tourist behaviour based on the tracking data collected by GPS. *Geographic information systems: Concepts, methodologies*. doi: 10.4018/978-1-4666-2038-4.ch066
- Palacios, Rafael. (2006). *Matlab Central*. Recuperado de <http://www.mathworks.com/matlab-central/fileexchange/10915-deg2utm>
- Shin, D., Aliaga, D., Tunçer, B., Müller, S., Kim, S., & Zünd, D. (2014). Urban sensing: Using smartphones for transportation mode classification. *EISEVIER*, 53. 76-86.
- Sohn, T., Varshavsky, A., LaMarca, A., Chen, M. Y., Choudhury, T., Smith, I., y otros. (2006). Mobility detection using everyday GSM traces. *UbiComp 2006: Ubiquitous computing*, 4206(221-224). Springer Berlin Heidelberg.
- Stenneth, L., Wolfson, O., Yu, P. S., & Xu, B. (November 2011). Transportation mode detection using mobile phones and GIS information. *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. Chicago, Illinois.
- Wu, L., Yang, B., & Jing, P. (2016). Travel mode detection based on GPS raw data collected by smartphones: a systematic review of the existing methodologies. *Multidisciplinary Digital Publishing Institute*, 7 (4), 67.
- Xiao, G., Juan, Z., & Zhang, C. (2015). Travel mode detection based on GPS track data and Bayesian networks. *Computers, environment and urban systems*, 54, 14-22.
- Yan, Z., Chakraborty, D., Parent, C., & Spaccaprieta, S. (2013). Semantic Trajectories: Mobility data computation and annotation. *ACM Transactions on Intelligent Systems and Technology*, 4 (3).
- Zhan , X., Hasan , S., Ukkusuri , S., & Kamga , C. (2013). Urban link travel time estimation using large-scale taxi data with partial information . *EISEVIER*, 33,37-49.
- Zhang, J., Arentze, T., & Timmermans, H. (2012). A multimodal transport network model for advanced traveler information system. *Journal of Ubiquitous Systems & Pervasive Networks*, 4 (1), 21-27.
- Zheng, Y., Chen, Y., Quanan, L., Xie, X., & Wei-Ying, M. (2010). Understanding Transportation Modes Based on GPS Data for Web Applications. *ACM*, 4 (1), 1-33.
- Zheng, Y., Liu, L., & Wang, L. (April 2008). Learning transportation mode from raw

GPS data for geographic applications on the web. *International World Wide Web Conference Committee (IW3C2)*. Beijing, China.

Zhu, X., Li, J., Liu, Z., Wang, S., & Yang, F. (June 2016). Learning Transportation Annotated Mobility Profiles from GPS Data for Context-Aware Mobile Services. *Services Computing (SCC), 2016 IEEE International Conference*.