

The logo of the Universidad Autónoma de México (UAM) is located in the top left corner of the page. It is a small, square icon with a black border. Inside the square, the text "img/logoUAM.png" is written in a small, black, sans-serif font.

img/logoUAM.png

Caracterización de datos de trayectorias individuales

Presentado por:
Jorge Rafael Martínez Buenrostro

Asesora: Dra. Elizabeth Pérez Cortés

México, CDMX, a 5 de diciembre de 2025

Contenido

Lista de Códigos	III
1. Introducción del Proyecto	1
1.1. Importancia de los modelos de movilidad para la evaluación de protocolos para redes móviles	1
1.2. Proceso de diseño de un modelo de movilidad	1
1.3. Objetivo del proyecto	1
1.4. Logros	2
2. Marco teórico	3
2.1. Elementos de un modelo de movilidad	3
2.1.1. Punto de recorrido	3
2.1.2. Tiempos de pausa	3
2.1.3. Longitud de vuelo	4
3. Objetivos y metodología	5
3.1. Objetivos	5
3.2. Metodología	5
4. Desarrollo	7
4.1. Caracterización inicial	7
4.2. Limpieza de datos	7
4.3. Determinación de puntos de recorrido	7
4.4. Determinación de tiempos de pausa	8
4.5. Determinación de longitudes de vuelo	8
5. Resultados	9
6. Conclusiones y Trabajo futuro	10

Lista de Figuras

Lista de Códigos

Capítulo 1

Introducción del Proyecto

1.1 Importancia de los modelos de movilidad para la evaluación de protocolos para redes móviles

La simulación de una red de comunicaciones en donde intervienen dispositivos personales de comunicación requiere de modelos que representen fielmente los patrones de movimiento de las personas. De lo contrario, las conclusiones derivadas de dicha simulación pueden ser poco útiles. Para avanzar hacia la definición de un modelo de trayectorias individuales, se propone caracterizar los datos de una base de datos existente que permita modelar trayectorias de una forma eficaz.

1.2 Proceso de diseño de un modelo de movilidad

El diseño de un modelo de movilidad implica varias etapas:

- **Caracterización de datos:** Limpieza, depuración y análisis exploratorio.
- **Identificación de trayectorias:** Extracción de secuencias de movimiento significativas.
- **Validación y evaluación:** Medición de la calidad y representatividad de las trayectorias.

1.3 Objetivo del proyecto

El objetivo principal del proyecto es obtener una caracterización estadística de las trayectorias individuales a partir de un conjunto de datos de movilidad. Esto incluye la identificación de trayectorias peatonales y su análisis mediante herramientas de IA.

1.4 Logros

- Reducir el conjunto de datos de 69.98 millones de registros, 19 columnas y un peso de 22 GB a 51 millones de registros, 7 columnas y 7 GB.
- Identificar que el 68.73 % de los registros tienen precisión GPS satelital (1-20 metros).
- Desarrollar un algoritmo de evaluación de calidad de trayectorias con métrica como volumen, cobertura temporal, precisión GPS y diversidad espacial.
- Encontrar un respaldo del algoritmo de evaluación usando el algoritmo de clusterización K Medias con los puntos de recorrido.
- Graficar la distribución de las longitudes de vuelo para poder aproximar una distribución probabilística.

Capítulo 2

Marco teórico

2.1 Elementos de un modelo de movilidad

Un modelo de movilidad describe el desplazamiento de individuos o grupos en el espacio y el tiempo. Este proyecto se enfoca en trayectorias peatonales individuales, definidas por:

- **Puntos de recorrido:** Ubicaciones específicas por las que transita el individuo.
- **Tiempos de pausa:** Intervalos durante los cuales el individuo permanece detenido en un mismo punto de recorrido.
- **Longitud de vuelo:** Distancia entre dos puntos de recorrido consecutivos.

A continuación se muestran la determinación de cada uno de estos elementos.

2.1.1 Punto de recorrido

Cada punto de recorrido se determinó a partir de los registros GPS contenidos en el conjunto de datos, específicamente mediante las coordenadas geográficas *device_lat* y *device_lon*. Para garantizar la calidad de estos puntos, se aplicó un filtro de precisión GPS, considerando únicamente aquellos registros con *device_horizontal_accuracy* menor a 10 metros, correspondiente a la precisión satelital.

2.1.2 Tiempos de pausa

Para determinar estos tiempos de pausa, se analizó la secuencia temporal de cada individuo. Cuando dos registros consecutivos presentaban coordenadas idénticas (o variaciones menores a un umbral de 0.001 grados, aproximadamente 111 metros), se interpretó que el individuo permaneció en pausa entre dichos registros. Adicionalmente, como no es posible saber en que parte del intervalo se encuentra el primer punto de recorrido registrado. Para calcular los tiempos de pausa se asume

que el primer punto de recorrido es cuando el individuo llegó a dicho punto. Esto aunque es una simplificación es de mucha utilidad para los objetivos del proyecto.

2.1.3 Longitud de vuelo

Para calcular la distancia entre dos puntos de recorrido, se utiliza la fórmula de Haversine aplicada a las coordenadas geográficas de registros sucesivos de un mismo individuo. Solo se consideraron desplazamientos significativos, definidos como aquellos donde el cambio de coordenadas superó un umbral de 0.001 grados (aproximadamente 111 metros). Este filtro permite distinguir entre movimientos reales y variaciones menores debidas al error de medición del GPS.

Capítulo 3

Objetivos y metodología

3.1 Objetivos

El objetivo principal del proyecto es obtener una caracterización estadística de las trayectorias individuales a partir de un conjunto de datos de movilidad.

Los objetivos particulares son:

- Caracterizar la base de datos para extraer las trayectorias contenidas.
- Aplicar un modelo de inteligencia artificial para identificar y analizar dichas trayectorias.

3.2 Metodología

El proceso se dividió en las siguientes etapas:

1. Caracterización inicial

- Exploración del conjunto de datos, identificación de columnas y filas irrelevantes.
- Conteo de registros y dimensiones de la base de datos.
- Inspección de valores únicos y estructura general.

2. Limpieza de datos

- Selección de campos relevantes para el análisis.
- Eliminación de columnas redundantes o sin valor analítico.
- Conservación de columnas clave: *identificador*, *timestamp*, *coordenadas*, *precisión GPS*.

- Eliminación de registros duplicados basados en *identificador*, *timestamp* y *coordenadas*.
- Filtrado de registros con baja precisión GPS mayor a 20 metros.
- Análisis de frecuencia de aparición de identificadores únicos.
- Desarrollo de algoritmo de puntuación compuesta basado en métricas
- Clasificación de trayectorias en categorías cualitativas.

3. Determinación de puntos de recorrido

- Distribución de individuos por día mediante histogramas temporales.
- Cálculo de frecuencia de aparición y persistencia temporal.
- Identificación de individuos con registros en múltiples días.
- Cálculo de velocidades entre puntos consecutivos.
- Creación de histograma de distribución de velocidades.
- Aplicación de filtro peatonal utilizando parámetros estadísticos:
 - Media de velocidad peatonal: 1.34 m/s
 - Desviación estándar: 0.37 m/s
 - Rango aceptable: 0.6-2.08 m/s (media \pm desviaciones estándar)
- Segmentación de trayectorias para eliminar puntos fuera del rango peatonal
- Mapa de distribución de puntos de recorrido por ciudad.
- Visualización de trayectorias individuales completas.
- Implementación de algoritmo de *clustering* sobre puntos de velocidad peatonal.

4. Determinación de tiempos de pausa

- Cálculo de tiempos de pausa.

5. Determinación de longitudes de vuelo

- Cálculo de longitud de vuelo entre puntos consecutivos.

4.1 Caracterización inicial

- Exploración del conjunto de datos, identificación de columnas y filas irrelevantes.
- Conteo de registros y dimensiones de la base de datos.
- Inspección de valores únicos y estructura general.

4.2 Limpieza de datos

- Selección de campos relevantes para el análisis.
- Eliminación de columnas redundantes o sin valor analítico.
- Conservación de columnas clave: *identificador*, *timestamp*, *coordenadas*, *precisión GPS*.
- Eliminación de registros duplicados basados en *identificador*, *timestamp* y *coordenadas*.
- Filtrado de registros con baja precisión GPS mayor a 20 metros.
- Análisis de frecuencia de aparición de identificadores únicos.
- Desarrollo de algoritmo de puntuación compuesta basado en métricas
- Clasificación de trayectorias en categorías cualitativas.

4.3 Determinación de puntos de recorrido

- Distribución de individuos por día mediante histogramas temporales.
- Cálculo de frecuencia de aparición y persistencia temporal.
- Identificación de individuos con registros en múltiples días.

- Cálculo de velocidades entre puntos consecutivos.
- Creación de histograma de distribución de velocidades.
- Aplicación de filtro peatonal utilizando parámetros estadísticos:
 - Media de velocidad peatonal: 1.34 m/s
 - Desviación estándar: 0.37 m/s
 - Rango aceptable: 0.6-2.08 m/s (media \pm desviaciones estándar)
- Segmentación de trayectorias para eliminar puntos fuera del rango peatonal
- Mapa de distribución de puntos de recorrido por ciudad.
- Visualización de trayectorias individuales completas.
- Implementación de algoritmo de *clustering* sobre puntos de velocidad peatonal.

4.4 Determinación de tiempos de pausa

- Cálculo de tiempos de pausa.

4.5 Determinación de longitudes de vuelo

- Cálculo de longitud de vuelo entre puntos consecutivos.

Capítulo 5

Resultados

Capítulo 6

Conclusiones y Trabajo futuro

Capítulo 6

Referencias

- [1] Autor referencia 1
- [2] Autor referencia 2
- [3] Autor referencia 3