

# GPS Trajectory segmentation and clustering method

Gary Reyes Zambrano<sup>1</sup>, Francisco Córdova Rizo<sup>1</sup>, Oscar León Granizo<sup>1</sup>, Edwin Carabali Noriega<sup>1</sup>

<sup>1</sup>Facultad de Ciencias Matemáticas y Físicas, Universidad de Guayaquil, Cda. Universitaria Salvador Allende, Guayaquil 090514, Ecuador, [gary.reyesz@ug.edu.ec](mailto:gary.reyesz@ug.edu.ec), [francisco.cordovar@ug.edu.ec](mailto:francisco.cordovar@ug.edu.ec), [oscar.leong@ug.edu.ec](mailto:oscar.leong@ug.edu.ec), [edwin.carabalin@ug.edu.ec](mailto:edwin.carabalin@ug.edu.ec)

*Abstract. - The analysis of GPS trajectories, in the context of current life, constitutes one of the main sources of support for decision making. The detection of patterns in GPS trajectories that allow determining actions to follow and decisions to be taken more quickly is one of the main lines of research in this field. In this paper we propose a method of segmentation and clustering of GPS trajectories composed of a segmentation algorithm that aims to segment the trajectories based on a partitioning criterion. Then the generated segments are used by a trajectory clustering algorithm based on traditional k-means; the method improves the clustering centroid selection process by using an initial grid and proposes a similarity function between trajectory segments for the clustering process. The results of the experiments performed show a significant improvement compared to the methods published in the literature.*

*Keywords: segmentation, clustering, GPS trajectories, intelligent transportation systems.*

**Digital Object Identifier (DOI):**

<http://dx.doi.org/10.18687/LACCEI2022.1.1.225>

ISBN: 978-628-95207-0-5 ISSN: 2414-6390

# Método de segmentación y agrupamiento de Trayectorias GPS

Gary Reyes Zambrano<sup>1</sup>, Francisco Córdova Rizo<sup>1</sup>, Oscar León Granizo<sup>1</sup>, Edwin Carabali Noriega<sup>1</sup>

<sup>1</sup>Facultad de Ciencias Matemáticas y Físicas, Universidad de Guayaquil, Cda. Universitaria Salvador Allende, Guayaquil 090514, Ecuador, [gary.reyesz@ug.edu.ec](mailto:gary.reyesz@ug.edu.ec), [francisco.cordovar@ug.edu.ec](mailto:francisco.cordovar@ug.edu.ec), [oscar.leong@ug.edu.ec](mailto:oscar.leong@ug.edu.ec), [edwin.carabalin@ug.edu.ec](mailto:edwin.carabalin@ug.edu.ec)

**Resumen:** El análisis de trayectorias GPS, en el contexto de la vida actual, constituye una de las principales fuentes de apoyo a la toma de decisiones. La detección de patrones en trayectorias GPS que permitan determinar acciones a seguir y decisiones que tomar con mayor rapidez es una de las principales líneas de investigación en este campo. En el presente trabajo se propone un método de segmentación y agrupamiento de trayectorias GPS compuesto por un algoritmo de segmentación que tiene como finalidad segmentar las trayectorias basado en un criterio de particionamiento. Luego los segmentos que se generan son utilizados por un algoritmo de agrupamiento de trayectorias basado en el tradicional k-means; el método mejora el proceso de selección de centroides del agrupamiento utilizando una cuadrícula inicial y propone una función de similitud entre segmentos de trayectorias para el proceso de agrupamiento. Los resultados de los experimentos realizados muestran una mejoría significativa en comparación con los métodos publicados en la bibliografía.

**Palabras clave:** segmentación, agrupación, trayectorias GPS, sistemas inteligentes de transporte.

## I. INTRODUCCIÓN

La creciente utilización de dispositivos GPS y la evolución en el campo de la transportación demandan de técnicas más eficientes para el análisis de datos y la toma de decisiones. Los sistemas inteligentes de transportación procesan grandes cúmulos de datos de trayectorias GPS generadas a partir de los vehículos que circulan por las vías en tiempo real [1], [2] Los datos recolectados deben ser analizados para convertirlos en conocimiento y utilizarlos como datos de apoyo a la toma de decisiones en proyectos tales como urbanismo o construcción de carreteras o conductos viales. Como parte del procesamiento de los datos en sistemas inteligentes de transportación se utilizan algoritmos para la agrupación de las trayectorias GPS basada en diferentes criterios [3], [4]. En la bibliografía se abordan varios métodos que realizan el agrupamiento basado en la segmentación de los datos y el cálculo de la similitud de estos segmentos, sin embargo, para la obtención de datos GPS que permitan realizar el proceso de extracción de conocimiento basado en elementos específicos como la velocidad, el tiempo, la dirección de las trayectorias estos algoritmos no presentan buenos resultados.

La principal razón reside en que realizan el análisis y agrupamiento de los datos teniendo en cuenta una función de

similitud que solo aborda la distancia entre dos puntos o entre dos segmentos, lo que limita el análisis de otros elementos de interés. Para utilizarlos en aplicaciones que requieran agrupar datos teniendo en cuenta varias variables es necesario realizar varias ejecuciones y transformaciones adicionales.

En la presente investigación se realiza un análisis de los algoritmos de agrupación de trayectorias GPS en la sección trabajos previos, se propone un método de segmentación y agrupamiento de trayectorias GPS en la sección método para la segmentación y agrupamiento de trayectorias GPS, se realizan varios experimentos para mostrar las fortalezas del método en la sección análisis y discusión y finalmente se concluye el trabajo y se proponen las directrices futuras para la continuación de la investigación.

## II. TRABAJOS PREVIOS

El agrupamiento de trayectorias GPS constituye un proceso de gran importancia. La clasificación de trayectorias basado en uno o más clasificadores y su agrupamiento en clústeres facilita el análisis. Como parte del presente trabajo se realiza un análisis documental asociado a los algoritmos de agrupamiento de trayectorias GPS más relevantes en la bibliografía.

Las técnicas de clustering han sido utilizadas en el análisis de trayectorias desde hace varios años. Por lo general, se trata de adaptaciones de los algoritmos convencionales utilizando métricas de similitud especialmente diseñadas para trayectorias [5], [6].

El algoritmo de agrupamiento Tra-DBscan [7] constituye una mejora del algoritmo clásico de agrupamiento DBscan para el análisis de trayectorias GPS. Este algoritmo utiliza la distancia de Hausdorff como métrica de similitud para calcular en qué medida los elementos (datos GPS) presentes en cada uno de los clústeres son similares entre sí. La métrica de distancia utilizada está compuesta por:

- La distancia perpendicular
- La distancia paralela
- La distancia angular

La diferencia de este algoritmo con respecto a DBScan consiste en la ejecución de un proceso de segmentación para realizar el análisis de segmentos de trayectorias GPS y la utilización de la métrica de distancia de Hausdorff como métrica de similitud entre elementos.

El agrupamiento de trayectorias posee ciertas características que deben ser consideradas especialmente. En primer lugar, es preciso definir la manera en que se ingresarán los datos que componen la trayectoria. Hacerlo de manera individual, procesando las ubicaciones una a una, dificulta la identificación de recorridos similares por lo que suele segmentarse previamente; tal es el caso de [8]–[10] que proponen utilizar segmentación en una etapa previa al clustering logrando mejorar el rendimiento de los algoritmos y una mejor calidad en los resultados del clustering.

En [11] se propone un modelo de trayectoria mejorado y se presenta un nuevo algoritmo de agrupación, con una medida de similitud que calcula la distancia entre dos trayectorias basado en múltiples características de los datos, logrando maximizar la similitud entre ellas. Por otra parte, en [12] se presenta una nueva técnica de agrupamiento de trayectorias que utiliza campos vectoriales para representar los centros de los grupos y proponer una definición de similitud entre trayectorias. Hoy en día continúan los esfuerzos de investigación en esta área como se evidencia en [13], [14].

TRACCLUS es un algoritmo de agrupamiento de trayectorias GPS basado en el marco de partición [9]. Su funcionamiento se basa en el análisis de segmentos pertenecientes a trayectorias GPS para determinar el nivel de similitud o disimilitud que presentan. El algoritmo define que:

Dados un conjunto de trayectorias representadas de la forma  $T = \{TR_1 \dots TR_n\}$ , siendo  $n$  el número de trayectorias a analizar, se genera un conjunto de clústeres  $C = \{C_1 \dots C_p\}$ , siendo  $p$  el número de clústeres.

Este algoritmo utiliza una medida de distancia compuesta por tres componentes:

1. Componente de distancia perpendicular.
2. Componente de distancia paralela.
3. Componente de distancia angular.

Las métricas de distancia se muestran en la figura 1. El algoritmo comparte varias características con el algoritmo DBSCAN, sin embargo, a diferencia de DBSCAN, no todos los conjuntos conectados por densidad pueden convertirse en conglomerados.

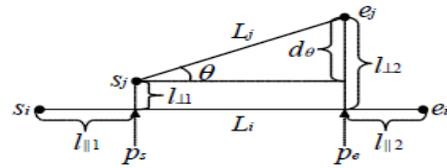


Figura 1. Métricas de distancia

$$d_{\perp} = \frac{l_{\perp 1}^2 + l_{\perp 2}^2}{l_{\perp 1} + l_{\perp 2}} \quad (1)$$

$$d_{||} = \text{MIN}(l_{||1}, l_{||2}) \quad (2)$$

$$d_{\theta} = |l_j| \times \sin(\theta) \quad (3)$$

En las ecuaciones 1, 2 y 3 se observan las formulaciones para obtener las tres medidas de distancias: perpendicular, paralela y angular respectivamente para tráculos.

En [10] se propone el algoritmo ATCGD (Adaptive Trajectory Clustering approach based on Grid and Density, por sus siglas en inglés) el cual mejora en un 95 por ciento al algoritmo TRACCLUS. Este algoritmo está compuesto por 3 fases denominadas partición, mapeo y agrupamiento. En la fase de partición, se aplica el método de partición MDL (AD-MDL) basado en la diferencia angular promedio de cada segmento. En la segunda fase se realiza el mapeo de los segmentos de trayectorias en las celdas correspondientes y en la fase de agrupamiento se utiliza un algoritmo basado en DBscan para agrupar los segmentos. Este algoritmo de agrupamiento se compara con el algoritmo TRACCLUS y los resultados se pueden observar en la tabla 1.

Tabla 1. Comparación entre los algoritmos analizados

Observaciones	TRACCLUS		ATCGD	
	Mediciones	Tiempo de ejecución	Mediciones	Tiempo de ejecución
HT-100	1486875	1,2500	1140856	0,1400
HT-200	5416222	5,8400	4327626	0,2300
HT-300	8164510	15,7500	7602455	0,4400
HT-400	9741195	26,3400	10682513	0,6100
RT1	461437	1,0700	39426	0,0900
RT2	164351	21,7500	176269	0,5700

A partir de esta comparación en la presente investigación se decide que las métricas utilizadas para el cálculo de la similitud de segmentos en el algoritmo ATCGD serán analizadas en profundidad con el objetivo de evolucionar la métrica de similitud propuesta en la sección 3.

En [15] se formula un algoritmo para la detección de puntos de subida y bajada de personas en taxis. Para la etapa de agrupamiento del algoritmo se utiliza el algoritmo DBscan

modificado y adaptado para trabajar con este tipo de datos en específico, denominado GADBSCAN. Los parámetros críticos de este algoritmo son el radio representado como  $d$  y la cantidad de puntos mínimos representados como *minptos*. En este caso el parámetro utilizado para la métrica de distancia es  $d$ , mientras que *minptos* es utilizado para delimitar la cantidad de puntos mínimos que aparecen en la vecindad al instante de asignar el punto y su vecindad a un clúster.

En [16] se propone un algoritmo para la agrupación de trayectorias GPS utilizando el algoritmo Fuzzy C-Means. Este algoritmo realiza el particionado de las trayectorias GPS utilizando un método basado en el ángulo de las trayectorias. Este algoritmo mejora el proceso de particionado de las trayectorias en segmentos de línea al utilizar un enfoque basado en el ángulo. La metodología descrita en este trabajo consiste en:

- Se utilizan los métodos de segmentación por ángulo y restricción basado en el coseno para generar segmentos de línea.
- Se utiliza K-means++ basado en Hausdorff para producir centroides iniciales, y se presenta un método basado en Lagrange para mejorar la agrupación de FCM.
- Se emplea el método de regresión de mínimos cuadrados para realizar la agrupación por regresión de trayectoria

Basado en el análisis documental realizado se considera que el uso de segmentos para el análisis de trayectorias GPS en vez de trayectorias completas brinda mejores resultados. Para el proceso de segmentación se puede constatar que los mejores resultados se encuentran los que utilizan el ángulo de cada punto. Otro elemento utilizado para la componente métrica de similitud es la distancia de Hausdorff. Esta métrica de distancia define el grado de similitud entre dos segmentos de trayectorias GPS lo cual es utilizado para asignar un segmento a un clúster en particular. Esta métrica es utilizada para la construcción de una nueva métrica de distancia que permita construir clústeres más robustos y eficientes.

#### A) Funciones de similitud

Las funciones de similitud se utilizan para evaluar la similitud entre elementos presentes en las trayectorias GPS. La clase de función de similitud elegida, por ejemplo, basadas en la distancia, en estadísticas o en puntos, depende del tipo de objeto y contexto considerado. En el caso de trayectorias GPS, la función debe tener en cuenta el gráfico subyacente de la red de carreteras y la conectividad de los gráficos o el cumplimiento del orden en las secuencias[17].

Entre las funciones de similitud más utilizadas en la literatura [18] se encuentran:

1. Distancia limitada por la red.
2. Distancia basada en la forma.
3. Distancia basada en la deformación (Warping)

De ellas para la presente investigación son de interés la distancia basada en la forma y la distancia basada en la deformación (warping). La distancia basada en la forma solo analiza la forma de la trayectoria mientras que la distancia basada en la deformación analiza también la componente temporal de las trayectorias GPS. Entre las métricas basadas en la deformación se encuentran la métrica de distancia euclidiana, la métrica de distancia manhattan y la métrica euclidiana síncrona. Entre las principales limitaciones de estas métricas se destacan en[18]:

- La función de coste  $\delta$  utiliza la distancia euclidiana. Algunas de estas distancias se han definido utilizando una norma L1, pero la distancia euclídea es más adecuada para los valores reales.
- Los métodos de deformación se basan en la comparación uno a uno entre secuencias. Por lo tanto, a menudo se requiere la elección de una serie particular que se utilizará como referencia, en la que todas las demás secuencias serán emparejadas. El indexado de dos secuencias que se comparan debe estar bien equilibrado para capturar mejor la variabilidad.
- El rendimiento de los métodos habituales basados en técnicas de deformación se ve obstaculizado por la gran cantidad de ruido inherentes a los datos sobre tráfico, lo que no es el caso cuando se estudian las series temporales.

Entre las métricas basadas en la forma se encuentra la distancia de Hausdorff y la distancia de Fréchet. La métrica de distancia de Hausdorff mide la distancia entre dos conjuntos. Para cada punto de la serie 1, se calcula la distancia mínima desde este punto hasta cualquier otro punto de la serie 2. El mayor de todas estas distancias define la distancia de Hausdorff. En la presente investigación se decide utilizar la distancia de Hausdorff como base para la formulación de la métrica de distancia a utilizar por el método de agrupamiento.

### III. MÉTODO PARA LA SEGMENTACIÓN Y AGRUPAMIENTO DE TRAYECTORIAS GPS

Una trayectoria GPS ha sido definida como un conjunto de información posicional, ordenada a partir del tiempo, de objetos en movimiento[18]. El creciente uso de receptores GPS permite

recoger una gran cantidad de datos, que pueden ser analizados para extraer información relevante. Para ello existen un conjunto de algoritmos de agrupamiento que, a partir de la información colectada, detectan diferentes patrones.

En la bibliografía analizada en la presente investigación se ha detectado tres formas de realizar el proceso de agrupación de trayectorias GPS[19]. Estas forman son:

1. Agrupamiento a partir del análisis de los puntos de dos trayectorias GPS.
2. Agrupamiento a partir del análisis de segmentos de trayectorias GPS.
3. Agrupamiento a partir del análisis de trayectorias GPS.

El análisis de segmentos de trayectorias GPS ha mostrado mejores resultados para el agrupamiento de trayectorias, por ello en la presente investigación se propone un método de análisis de trayectorias basado en el método de agrupamiento clásico K-means que utiliza los segmentos de trayectorias que se obtienen a partir de una segmentación basada en ángulos, se propone una métrica de agrupamiento y se describe el proceso de agrupación. Entre las principales mejoras se encuentran:

- El proceso de segmentación, partición de las trayectorias en partes más pequeñas tomando como criterio un valor umbral (valor del ángulo) entre un punto y el ángulo que forma con el siguiente punto.
- Se realiza un análisis local y global de los datos, lo que mejora el proceso de selección de centroides y garantiza que los grupos tienen elementos semejantes entre sí y diferentes con otros grupos.
- El trazado de una malla que divide el espacio de trayectorias GPS en 16 cuadrículas. En cada cuadrícula se trazan centroides direccionales, de esta manera se analiza la dirección de la trayectoria y la similitud entre los segmentos que pertenecen a cada cuadrícula. Esto constituye el análisis local de los datos.
- El proceso de filtrado se realiza además por tiempo, teniendo en cuenta las trayectorias que confluyen en un momento de tiempo determinado, lo que posibilita la detección de varios patrones dentro del clúster, tales como el congestionamiento vehicular.
- Una nueva métrica de similitud de segmentos de trayectorias mejora el recálculo de los centroides y la consolidación de los datos en el análisis global de los mismos.

El análisis de datos GPS constituye uno de los tópicos de mayor interés en el campo de los sistemas inteligentes de transportación pues aportan elementos de apoyo a la toma de decisiones que favorecen el desarrollo e impactan en la sociedad. La detección de congestionamientos vehiculares, la detección de patrones anómalos en el tráfico que ayuden a la predicción de accidentes y la evaluación del rendimiento de caminos y avenidas principales son algunas de los principales escenarios de aplicación. El método de segmentación y agrupamiento de trayectorias GPS que se propone en la presente investigación consta de dos componentes:

- A. Segmentación de trayectorias GPS
- B. Agrupamiento de sub-trayectorias GPS

El componente de segmentación de trayectorias GPS realiza la subdivisión de las trayectorias en sub-trayectorias a partir del análisis de los ángulos que forman los puntos que las componen. El componente de agrupamiento de trayectorias GPS realiza el agrupamiento de las trayectorias basado en un criterio de similitud.

#### A. Segmentación de trayectorias GPS

Este componente es el primero del método que se propone en la presente investigación. Su funcionamiento se basa en recorrer toda la trayectoria, punto a punto, buscando lugares para delimitar y convertir los puntos analizados en sub-trayectorias o segmentos. Para este proceso se utiliza el teorema del coseno en el cálculo de los ángulos que se encuentran entre dos puntos.

El proceso de segmentación inicia en el primer punto de la trayectoria y realiza el cálculo del ángulo entre los puntos denominados  $P_i$  y  $P_{i+1}$ . Este resultado se toma como referencia para comparar el ángulo resultante entre los puntos  $P_{i+1}, P_{i+2}$ . Si la diferencia entre ellos supera el valor umbral definido entonces se crea una sub-trayectoria. Este proceso se repite hasta terminar la trayectoria, analizando los  $n$  puntos de la trayectoria y formando  $x$  cantidad de segmentos o sub-trayectorias GPS. Estos segmentos pasan a ser la entrada del método de agrupamiento de sub-trayectorias GPS.

```

programa Segmentación (listapuntos: lista)
begin
  listaresultante: lista
  segmento: lista
  x: int
  angle = Angulo puntos (listapuntos [x],
  listapuntos [x+1])
  x = x + 1
  Mientras (¡listapuntos! = null) hacer
    Si Angulo puntos (listapuntos [x],

```

```

listapuntos[x+1]) < umbral
segmento. Adicionar(listapuntos[x])
x = x + 1
sino
lista resultante. adicionar(segmento)
x = x + 1
segmento. limpiar ()
fin Si
fin Mientras
fin.

```

### B. Agrupamiento de sub-trayectorias GPS

El agrupamiento de segmentos de trayectorias GPS es fundamental para la detección de patrones en trayectorias GPS. Como parte del método de agrupamiento de sub-trayectorias GPS se encuentra el diseño de una métrica de similitud de segmentos de trayectorias la cual determina los centroides y elementos que deben estar asociados a este. La formulación está dada por la métrica de similitud que se presenta en la ecuación 4.

$$D(T_1, T_2) = \frac{|(T_{1i} - T_{2f}) + (T_{1f} - T_{2i})|}{2} + D_a \quad (4)$$

Donde T representa el punto inicial y final de cada una de las trayectorias. El concepto en esta idea es calcular la distancia entre dos segmentos teniendo en cuenta cuanto es el espacio que existe entre ambas y el ángulo. Para el cálculo de  $D_a$  se propone la ecuación 5

$$D_a(T_1, T_2) = 1/n \sum_{i=0}^n \arccos(P_1 - P_2) \quad (5)$$

Una vez determinada la similitud de dos sub-trayectorias se realiza la agrupación propiamente dicha, momento en el cual se adiciona la sub-trayectoria al grupo al cual pertenece.

## IV. ANÁLISIS Y DISCUSIÓN

El agrupamiento de trayectorias GPS brinda la posibilidad de descubrir conocimiento en un dominio específico, bajo parámetros deseados y apoyar el proceso de toma de decisiones.

Para validar el correcto funcionamiento del método segmentación y agrupamiento propuesto se realiza la ejecución de dos experimentos. Un primer experimento utilizando un Conjunto de datos que contiene registros de trayectorias GPS de taxis que circulan por la ciudad de Beijing-China[20] para observar el comportamiento de las variables distancia media al centroide y desviación estándar. Esto ayuda a conocer la eficacia de la técnica de reticulado para la creación de centroides y asignación de elementos a los centroides. El segundo experimento tiene como objetivo comparar los

resultados obtenidos con resultados publicados en la bibliografía utilizando los mismos datos.

Se realizó un primer experimento en el cual se ejecuta el método propuesto con 31 centroides y tolerancia angular = 5, sobre el conjunto de datos de taxis de Beijing-China. Esta base de datos tiene 9526 trayectorias GPS compuestas por 4512520 puntos. Para la asignación de los centroides se traza una malla que divide en 16 zonas a todo el espectro de trayectorias a analizar. Los centroides se definen en diferentes direcciones y se definen los intervalos de tiempo deseados para la agrupación.

Esta estrategia para la asignación de centroides constituye uno de los aportes de la presente investigación. Como salida se tienen los diferentes grupos formados por segmentos de trayectorias GPS los cuales se representan en la figura 2.

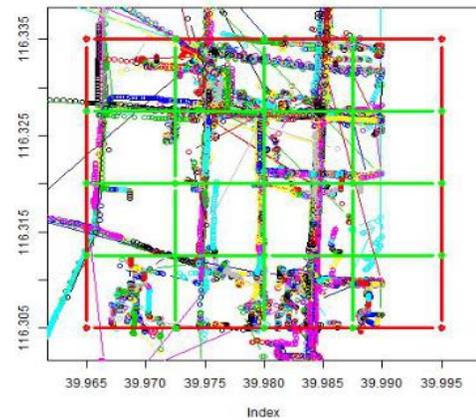


Figura 2. Representación de los resultados para  $k = 31$

En la tabla 2 se muestran los valores de mínimo, medio y máximo que fueron calculados a partir de los resultados obtenidos en los experimentos. Se calcula el valor de la métrica de Silhouette para el agrupamiento de 20 y 40 clústeres, obteniendo valores de 0.373902 y 0.401686 respectivamente.

Tabla 2. Valores resumidos del experimento uno

Métrica	Media al centroide	Desviación estándar
min	0.01031775	0.003741198
med	0.015315112	0.006830536
max	0.021662219	0.009600246

Con el objetivo de establecer una comparación con algunos de los algoritmos analizados se establece el segundo experimento. Este se diseña para seguir la misma metodología y utilizar los datos expuestos en [16] los cuales están compuestos por 27385 segmentos de línea y 71375 puntos. Se calcula la métrica *pbm-index* la cual aparece explicada y referenciada en [16]. En la tabla 3 se muestra una comparación

entre los resultados experimentales publicados y los obtenidos por el método propuesto.

Tabla 3 Comparación de datos experimentales de la bibliografía y propuestos en la presente investigación

Valores	K-means	Kmedias	FCM	FCML	Método propuesto
<b>K = 20</b>					
Máximo	0,0724	0,0764	0,0807	0,0903	0,4668
Media	0,0598	0,0615	0,0800	0,0868	0,1985
Mínimo	0,0457	0,0460	0,0791	0,0868	0,1201
<b>K = 40</b>					
Máximo	0,0535	0,0545	0,0620	0,0662	0,2334
Media	0,0477	0,0458	0,0607	0,0677	0,0753
Mínimo	0,0402	0,0279	0,0600	0,0671	0,0446

Como se puede apreciar en la tabla 3 los resultados obtenidos por la métrica *pbm-index* son mejores que los propuestos en la bibliografía. La razón principal de la mejoría de los resultados reside en la propuesta de métrica para calcular la similitud entre dos segmentos de trayectorias. Esta métrica es utilizada tanto para calcular la similitud de los elementos inter-clúster como intra-clúster en el proceso de recalcu de centroides.

#### CONCLUSIONES Y TRABAJO FUTURO

El método de agrupamiento de trayectorias GPS propuesto en la presente investigación realiza la segmentación de trayectorias GPS basado en ángulos y permite analizar las trayectorias segmentadas a partir de un valor umbral de ángulo determinado. La utilización de una nueva métrica de distancia que facilite medición espacial y temporal de los segmentos de trayectorias mejora el proceso de agrupamiento y facilita el descubrimiento de conocimiento en los datos minados. Con la aplicación del método de agrupación propuesto se pueden detectar varios patrones dentro de las trayectorias GPS.

Como trabajo futuro se pretende mejorar el método con diferentes actualizaciones, tanto en la función de similitud como en el algoritmo de segmentación. Se trabajará en la detección de diferentes patrones entre los que se encuentra la congestión vehicular.

REFERENCIAS

- [1] N. Marković, P. Sekuła, Z. vander Laan, G. Andrienko, and N. Andrienko, “Applications of Trajectory Data from the Perspective of a Road Transportation Agency: Literature Review and Maryland Case Study,” Aug. 2017, [Online]. Available: <http://arxiv.org/abs/1708.07193>
- [2] H. Menouar, I. Guvenc, K. Akkaya, A. S. Uluagac, A. Kadri, and A. Tuncer, “UAV-Enabled Intelligent Transportation Systems for the Smart City: Applications and Challenges,” 2017, Accessed: Feb. 13, 2022. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7876852>
- [3] Y. Wang, K. Qin, Y. Chen, and P. Zhao, “Detecting anomalous trajectories and behavior patterns using hierarchical clustering from Taxi GPS Data,” *ISPRS International Journal of Geo-Information*, vol. 7, no. 1, Jan. 2018, doi: 10.3390/ijgi7010025.
- [4] M. Fountoulakis, N. Bekiaris-Liberis, C. Roncoli, I. Papamichail, and M. Papageorgiou, “Highway traffic state estimation with mixed connected and conventional vehicles: Microscopic simulation-based testing,” *Transportation Research Part C: Emerging Technologies*, vol. 78, pp. 13–33, May 2017, doi: 10.1016/j.trc.2017.02.015.
- [5] M. Y. Choong, R. K. Y. Chin, K. B. Yeo, and K. T. K. Teo, “Trajectory pattern mining via clustering based on similarity function for transportation surveillance,” *International Journal of Simulation: Systems, Science and Technology*, vol. 17, no. 34, pp. 19.1-19.7, 2016, doi: 10.5013/IJSSST.a.17.34.19.
- [6] J. Kim and H. S. Mahmassani, “Spatial and temporal characterization of travel patterns in a traffic network using vehicle trajectories,” in *Transportation Research Procedia*, 2015, vol. 9, pp. 164–184. doi: 10.1016/j.trpro.2015.07.010.
- [7] L. Liu, J. Song, B. Guan, Z. Wu, and K. He, “Tra-DBScan: un algoritmo de trayectorias de agrupamiento”, Accessed: Feb. 13, 2022. [Online]. Available: <https://www.scientific.net/AMM.121-126.4875>
- [8] G. Yuan, P. Sun, J. Zhao, D. Li, and C. Wang, “Una revisión de los algoritmos de agrupamiento de trayectorias de objetos en movimiento,” 2016, Accessed: Feb. 13, 2022. [Online]. Available: <https://link.springer.com/article/10.1007/s10462-016-9477-7>
- [9] J.-G. Lee, J. Han, and K.-Y. Whang, “Trajectory Clustering: A Partition-and-Group Framework \*.”
- [10] Y. Mao, H. Zhong, H. Qi, P. Ping, and X. Li, “An adaptive trajectory clustering method based on grid and density in mobile pattern analysis,” *Sensors (Switzerland)*, vol. 17, no. 9, Sep. 2017, doi: 10.3390/s17092013.
- [11] Q. Yu, Y. Luo, C. Chen, and S. Chen, “Trajectory similarity clustering based on multi-feature distance measurement,” *Applied Intelligence*, vol. 49, no. 6, pp. 2315–2338, Jun. 2019, doi: 10.1007/s10489-018-1385-x.
- [12] N. Ferreira, J. T. Klosowski, C. Scheidegger, and C. Silva, “Vector Field k-Means: Clustering Trajectories by Fitting Multiple Vector Fields,” Aug. 2012, [Online]. Available: <http://arxiv.org/abs/1208.5801>
- [13] J. Lou and A. Cheng, “Detecting pattern changes in individual travel behavior from vehicle GPS/GNSS data,” *Sensors (Switzerland)*, vol. 20, no. 8, Apr. 2020, doi: 10.3390/s20082295.
- [14] H. Hu, G. Lee, J. H. Kim, and H. Shin, “Estimating micro-level on-road vehicle emissions using the k-means clustering method with GPS big data,” *Electronics (Switzerland)*, vol. 9, no. 12, pp. 1–18, Dec. 2020, doi: 10.3390/electronics9122151.
- [15] Y. Shen, L. Zhao, and J. Fan, “Analysis and visualization for hot spot based route recommendation using short dated taxi GPS traces,” *Information (Switzerland)*, vol. 6, no. 2, pp. 134–151, 2015, doi: 10.3390/info6020134.
- [16] X. Zhou, F. Miao, H. Ma, H. Zhang, and H. Gong, “A trajectory regression clustering technique combining a novel fuzzy C-means clustering algorithm with the least squares method,” *ISPRS International Journal of Geo-Information*, vol. 7, no. 5, May 2018, doi: 10.3390/ijgi7050164.
- [17] E. D’andrea, D. di Lorenzo, B. Lazerini, F. Marcelloni, and F. Schoen, “Path Clustering based on a Novel Dissimilarity Function for Ride-Sharing Recommenders.” [Online]. Available: [www.RideshareOnline.com](http://www.RideshareOnline.com)
- [18] P. Besse, B. Guillouet, J.-M. Loubes, and R. François, “Review and Perspective for Distance Based Trajectory Clustering,” Aug. 2015, [Online]. Available: <http://arxiv.org/abs/1508.04904>
- [19] P. C. Besse, B. Guillouet, J. M. Loubes, and F. Royer, “Review and Perspective for Distance-Based Clustering of Vehicle Trajectories,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 11. Institute of Electrical and Electronics Engineers Inc., pp. 3306–3317, Nov. 01, 2016. doi: 10.1109/TITS.2016.2547641.
- [20] J. Quemada and ACM Digital Library., *Proceedings of the 18th international conference on World wide web.* ACM, 2009.