

Visualización y análisis de trayectorias vehiculares con QGis y Weka

[Visualization and analysis of vehicle paths with QGis and Weka]

Gary Reyes Zambrano, Fátima Macías Tepán, Alexander Pizarro Jaime, and Jairo Macías Guevara

Facultad de Ciencias Matemáticas y Físicas,
Universidad de Guayaquil,
Guayaquil, Ecuador

Copyright © 2016 ISSR Journals. This is an open access article distributed under the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT: This article presents a visual analysis and processing of a data set of vehicle GPS paths for the discovery of spatiotemporal patterns, which is developed in the Weka tool and QGis, the proposed analysis is based on the location of length and presents latitude of a set of vehicles which determines how close are the north pole and likewise verify that industry make more geographic coordinates. To perform this experiment was taken as the first step in selecting a software environment and appropriate hardware, enabling configuration and correct implementation of applications and thus able to perform better in the development of research, the following steps They are to create a base of spatial data, load the spatial data set using a procedure developed in Visual Studio, generation of geometric points, display of geometric points loaded from the database within a geographic information system QGis and finally analyze and process GPS vehicle trajectories using the technique of data mining, which it was found that there are a number of vehicles that make it coordinates near north pole.

KEYWORDS: spatio-temporal, spatial database, geographic information system, data mining.

RESUMEN: En este artículo se presenta un análisis visual y procesamiento de un conjunto de datos de trayectorias GPS vehiculares para el descubrimiento de patrones espacio-temporales, que está desarrollado en la herramienta Weka y en QGis, el análisis propuesto se base en la ubicación de longitud y latitud de un conjunto de vehículos él cual permite determinar qué tan cerca se encuentran del polo norte y así mismo verificar en que sector marcan más coordenadas geográficas. Para poder realizar este experimento se tomó como primer paso la selección de un ambiente de software y hardware adecuado, que permitió la configuración y ejecución correcta de las aplicaciones y de esta manera poder tener un mejor desempeño en el desarrollo de la investigación, los siguientes pasos consisten en crear una base de dato espacial, cargar el conjunto de datos espaciales mediante un procedimiento desarrollado en Visual Studio, generación de puntos geométricos, visualización de los puntos geométricos cargados desde la base de datos dentro de un sistema de información geográfico QGis y finalmente analizar y procesar las trayectorias GPS vehiculares aplicando la técnica de minería de datos, lo cual se pudo comprobar que existen un número de vehículos que marcan coordenadas cerca de él polo norte.

PALABRAS CLAVES: espacio-temporales, base de datos espacial, sistema de información geográfico, minería de datos.

1 INTRODUCCION

En la actualidad el aumento considerable de información espacial en las organizaciones, hacen necesarias la explotación de esta por medio de la minería de datos, ya que por medio de esta técnica pueden ser procesados grandes volúmenes de

información con el objetivo de extraer patrones que residan en los datos almacenados. La minería de datos es una tecnología novedosa que integra diferentes técnicas de análisis de datos y extracción de modelos.

Actualmente las empresas generan grandes volúmenes de información, siendo este su activo más importante e indispensable para la toma de decisiones. Esto ha hecho necesario el desarrollo de herramientas que analicen y procesen los datos para identificar la información más relevante. Para el almacenamiento de la información se utilizan las bases de datos relacionales y Datawarehouse por sus características de capacidad de soporte, estrategia que ha dado muy buenos resultados en cuanto a la carga estructurada de datos, sin embargo el tiempo de respuesta en el procesamiento y análisis es solo aceptable. [1]

En los últimos años, no obstante, la etapa en que los datos básicamente se almacenan en bases de datos espaciales SIG para confeccionar cartografías o mapas puramente descriptivos, está evolucionando a una nueva fase caracterizada por un mayor peso del análisis numérico intensivo de los datos disponibles.[2]Por otra parte, la evolución de la informática y la electrónica ha dado al surgimiento de Sistemas de Información Geográfica, que constituyen una poderosa herramienta que considera componentes espaciales, temáticas y temporal de los datos y provee un sistema para la gestión de dichos datos georeferenciados, con herramientas de visualización, consulta, edición, consulta y análisis.

Es por esto que en el presente documento se tomó a consideración un conjunto de datos geográficos del proyecto (Microsoft Research Asia) Geolife de un total de 10 usuarios[3]. Este conjunto de datos está compuesto por trayectorias vehiculares que se detallan de la siguiente manera: longitud, latitud, altitud, tiempo y fecha en la que se generó la trayectoria.

En este documento vamos a describir el procesamiento de los datos, tomando a consideración los atributos de longitud y latitud de una trayectoria vehicular, aplicando filtros estos datos generaran intervalos donde se ubicaran cuáles son los vehículos que tienen una aproximación más cercana al punto de referencia y ver donde se han generado más coordenadas geográficas.

Los resultados podrán visualizarse en gráficos que la propia herramienta Weka proporciona.

2 INTELIGENCIA DE NEGOCIO ESPACIAL

La inteligencia de negocios espaciales permite tomar decisiones con base en el análisis de información espacial, combinando elementos de procesamiento analítico y sistemas de información geográfica para la solución de problemas espaciales[4].

COMPONENTES

Los componentes de un sistema de apoyo a las decisiones espaciales, hacen posible la extracción en interpretación de conocimiento, la tecnología de la información, datos espaciales, aplicaciones, usuarios, desarrolladores, interfaces de usuario, modelados explícitos y expertos [5].

Cada uno de estos componentes ejecutan unos pasos que permiten extraer, limpiar, transformar y organizar los datos y así obtener un consolidado de información espacial y alfanumérica, para que finalmente se llegue a tomar decisiones relacionadas con un problema específico. En la figura 1 muestra la estructura de la arquitectura en donde se encuentran los datos externos, gestión de datos, modelos, diálogos y conocimiento para producir las decisiones.

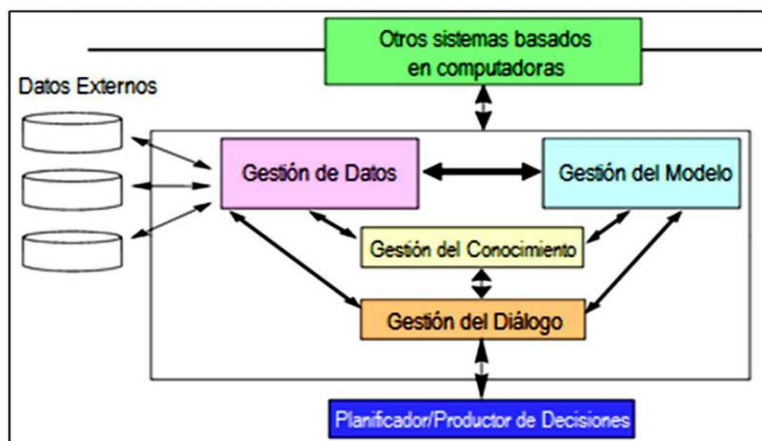


Fig. 1. Componentes de Decisiones Espaciales

Es una técnica de inteligencia de negocio que se ha venido utilizando, la cual a partir de la exploración y el análisis se enfoca en descubrir conocimiento[5]. Además la minería de datos es un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que encuentran ocultos en los datos.

Para poder llevar a cabo el descubrimiento de conocimiento se deben seguir una serie de pasos iterativos: limpieza, integración, selección, transformación, minería de datos, evaluación de patrones y representación de conocimientos.

La información espacial ha estado asociada en forma directa con la cartografía, en el logro de objetivos específicos concernientes con operaciones de análisis y gestión de datos espaciales, en las cuales se representan dicha información con modelos que utilizan mapas y símbolos[6].

TAREA DE LA MINERÍA DE DATOS

El proceso de minería involucra ajustar modelos o determinar patrones a partir de datos. Este ajuste normalmente es de tipo estadístico, en el sentido que se permite un cierto ruido o error dentro del modelo.[7]

Los algoritmos de minería de datos realizan en general tareas de predicción (de datos desconocidos) y de descripción (de patrones).

PROCESO DE LA MINERÍA DE DATOS

El proceso de minería de datos se inicia con la identificación de los datos[7]. Para ello hay que imaginar qué datos se necesitan, dónde se pueden encontrar (en una o varias bases de datos; en papel, etc.) y cómo conseguirlos. Una vez que se dispone de los datos, se deben preparar, poniéndolos en bases de datos en un formato adecuado o construir una warehouse. Esta es una de las tareas más difíciles minería de datos. Una vez que se tiene los datos en el formato adecuado hay que realizar una selección de los datos esenciales y eliminación de los innecesarios.

Antes de proceder al análisis de los datos, conviene tener una idea de qué es lo que interesa averiguar, qué herramientas se necesitan y cómo proceder. Tras aplicar la herramienta elegida o construida por nosotros mismos hay que saber interpretar los resultados o patrones obtenidos para saber los que son significativos y cómo poderlos para extraer únicamente los resultados útiles. Tras examinar los resultados útiles hay que identificar las acciones que deben de ser tomadas, discutir las y pensar en los procedimientos para llevarlas a cabo e implementarlas.

Una vez implementadas hay que evaluarlas para ello hay que observar los resultados, los beneficios y el coste para poder reevaluar el procedimiento completo. Para entonces los datos pueden haber cambiado, nuevas herramientas pueden estar disponibles y probablemente habrá que planificar el siguiente ciclo de minería.

3 DESARROLLO

TEGNOLOGIAS INVOLUCRADAS

Para el desarrollo de esta investigación se tomaron en cuenta los siguientes recursos:

- Laptop Dell latitude
- Características
- Windows 10 Enterprise
- Procesador Intel Core i5-250M CPU @ 2.50GHz
- Memoria de 8.00 GB
- Sistema Operativo de 64bits, procesador x64

En la tabla 1 se describe las aplicaciones que se utilizaron para realizar el procesamiento de los datos.

Tabla 1. Descripción de los programas

| Programas | Descripción |
|--------------------|---|
| Postgresql 9.3 | Sistema de administración de bases de datos que nos permite la gestión de datos.[8] |
| Postgis 9.3 | Es un módulo que añade soporte de objetos geográficos a la base de datos objeto-relacional PostgreSQL[9], convirtiéndola en una base de datos espacial para su utilización en Sistema de Información Geográfica. |
| QGis 2.6 | Sistema de Información geográfica que nos permite visualizar el mapa con sus respectivas trayectorias vehiculares. [10] |
| Weka 3.8 | Waikato Environment for Knowledge Analysis es una herramienta que permite la experimentación de análisis de datos mediante la aplicación, análisis y evaluación de las técnicas más relevantes de análisis de datos, principalmente las provenientes del aprendizaje automático, sobre cualquier conjunto de datos del usuario.[11] |
| Visual Studio 2015 | Es un conjunto de herramientas para crear software, desde la fase de diseño pasando por las fases de diseño de la interfaz de usuario, codificación, pruebas, depuración, análisis de la calidad y el rendimiento del código, implementación en los clientes y recopilación de telemetría de uso. |

PROCESAMIENTO DE DATOS

PostGIS es una extensión que convierte el sistema de base de datos PostgreSQL en una base de datos espacial, tomando en cuenta este concepto se creó una base de datos espacial con los siguientes campos que se los describe en la tabla 2.

Tabla 2. Campos de la tabla trayectoria

| Campos | Descripción |
|-----------|---|
| Id | Identificación del vehículo |
| latitude | Latitud: Es la distancia angular que hay desde un punto de la superficie de la Tierra. |
| longitude | Longitud: Es la Distancia angular de un punto de la superficie terrestre al meridiano de Greenwich. |
| altitude | Altitud: Es la distancia vertical de un origen dado, considerado como nivel cero. |
| date_tr | Fecha: Fecha de circulación |
| Time_tr | Hora: Hora de circulación |
| geom | Geométrico: almacena información geográfica. |

CARGA DE DATOS

El data set utilizado para esta investigación tiene un tamaño de 299 MB, debido al gran volumen de registros que almacena este conjunto de datos se realizó un procedimiento llamado “Cargar Geolocalización” en Visual Studio versión 2015, que permitió la importación de los registros y a su vez la creación del campo geométrico en el mismo procedimiento, para luego ser cargados todos los registros en la base de datos creada en postgis, este procedimiento tomo 60 minutos para la carga total de los registros. En la figura 2 muestra la interfaz gráfica del procedimiento cargar geolocalización.

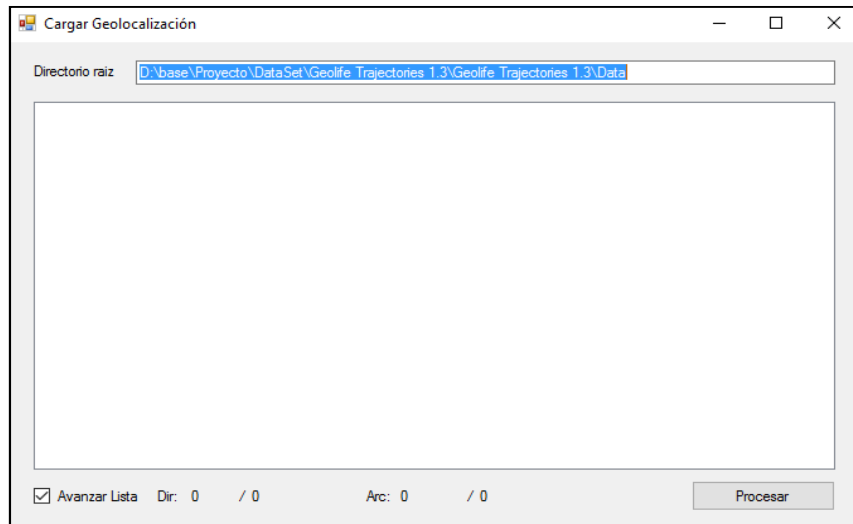


Fig. 2. Interfaz gráfica de Cargar de Geolocalización

VISUALIZACION DEL MAPA EN EL SISTEMA DE INFORMACIÓN GEOGRAFICA (QGIS)

En la figura 3, se muestra el mapa de China que fue descargado de OpenStreepMap que es una iniciativa a nivel mundial que tiene como objetivo el desarrollar una base cartográfica que esté libre de restricciones para su uso y distribución[12], este procedimiento tuvo una duración de 50 minutos aproximadamente.



Fig. 3. Mapa de China cargado en QGis

CARGAR TRAYECTORIAS VEHICULARES

Se realiza una conexión a la base de datos espacial desde QGis para de esta manera poder visualizar las trayectorias vehiculares en el mapa. En la figura 4 muestra el mapa con las capas que generan los puntos geométricos por cada vehículo, cada color que muestra en el mapa son los movimientos que genera cada vehículo.



Fig. 4. Mapa con las trayectorias vehiculares

4 RESULTADOS Y DISCUSION

El proceso y análisis de los datos se lo realiza mediante la herramienta Weka, ya que nos permite aplicar una gran diversidad de filtros y algoritmos sobre los datos permitiendo realizar transformaciones sobre ellos de todo tipo[13].

Para el desarrollo de este experimento vamos a tomar a consideración las trayectorias vehiculares de 10 vehículos, la cual analizaremos y posteriormente mostraremos los resultados y las conclusiones de nuestro experimento. Para cargar los datos fue necesario aumentar la memoria de weka a 4Gb con la finalidad de obtener los resultados de manera más rápida.

PREPROCESADOS

FILTROS

WEKA tiene integrados filtros que permiten realizar manipulaciones sobre los datos en dos niveles: atributos e instancias. Las operaciones de filtrado pueden aplicarse “en cascada”, de manera que cada filtro toma como entrada el conjunto de datos resultante de haber aplicado un filtro anterior[14].

Vamos a aplicar el filtro discretizar, este filtro es bastante útil cuando se trabaja con atributos numéricos[14], este filtrado nos permitirá dividir la amplitud total de intervalos, con diferentes opciones para seleccionar límites.

En la tabla 3 muestra la lista de atributos que van utilizado en el filtrado de datos.

Tabla 3. Atributos analizados

| N. | Atributos |
|----|----------------|
| 1 | latitude |
| 2 | longitude |
| 3 | Id_dispositivo |

Aplicando el filtro discretizar para los atributos mostramos en la figura 5 el atributo latitude y en la figura 6 el atributo longitude, donde nos indican la generación del intervalo en la columna label y en la columna count indica cuantos puntos pertenecen a cada intervalo.

Selected attribute

Name: latitudo
Missing: 0 (0%)
Distinct: 10
Type: Nominal
Unique: 0 (0%)

| No. | Label | Count | Weight |
|-----|-------------------------|---------|-----------|
| 1 | '(-inf-21.005122]' | 15898 | 15898.0 |
| 2 | '(21.005122-23.755607]' | 9689 | 9689.0 |
| 3 | '(23.755607-26.506092]' | 17404 | 17404.0 |
| 4 | '(26.506092-29.256576]' | 14574 | 14574.0 |
| 5 | '(29.256576-32.007061]' | 138887 | 138887.0 |
| 6 | '(32.007061-34.757545]' | 160574 | 160574.0 |
| 7 | '(34.757545-37.50803]' | 219201 | 219201.0 |
| 8 | '(37.50803-40.258514]' | 2124425 | 2124425.0 |

Fig. 5. Generación de intervalos al atributo latitud

Selected attribute

Name: longitudo
Missing: 0 (0%)
Distinct: 10
Type: Nominal
Unique: 0 (0%)

| No. | Label | Count | Weight |
|-----|---------------------------|---------|-----------|
| 1 | '(-inf-81.34267]' | 2018 | 2018.0 |
| 2 | '(81.34267-86.705036]' | 23478 | 23478.0 |
| 3 | '(86.705036-92.067402]' | 65650 | 65650.0 |
| 4 | '(92.067402-97.429767]' | 65712 | 65712.0 |
| 5 | '(97.429767-102.792133]' | 83381 | 83381.0 |
| 6 | '(102.792133-108.154498]' | 24411 | 24411.0 |
| 7 | '(108.154498-113.516863]' | 113771 | 113771.0 |
| 8 | '(113.516863-118.879229]' | 2251980 | 2251980.0 |

Fig. 6. Generación de intervalos al atributo longitudo

CLASIFICACION

Es la habilidad para adquirir una función que mapee un elemento de dato a una de varias predefinidas[7]. El problema de la clasificación es el más frecuente en la práctica. En este caso aplicaremos un algoritmo J48 que se encuentra dentro del grupo de árboles (tree).

ALGORITMO J48

El algoritmo J48 nos permite establecer una relación entre diversos tipos de información aplicando cálculos y heurística para el análisis de datos[15].

De esta forma construiremos un modelo que permita predecir la categoría de las instancias en función de los atributos longitud.

En la información que se despliega dentro de este algoritmo están.

- Instancias: 2781512
- Atributos: 3

El análisis de este algoritmo en relación al atributo longitud, lo explicaremos en la Figura 7, la interpretación de la primera línea es la siguiente. Si el atributo latitudo está en un intervalo de -inf-21.005122 y la longitudo en un intervalo de 108.154498 – 113.516863 entonces tenemos 15898 puntos de los 10 diferentes vehículos que estamos evaluando. En la segunda línea encontramos que si la latitudo está en un intervalo de 21.005122 – 23.755607 y dentro de este intervalo evaluamos los dispositivos que son menores o igual (tercera línea) y mayor que 7(cuarta línea), cada uno de las condiciones

tienen un intervalo de longitud que muestran puntos que marcados dependiendo si son mayores o menores que 7. De esta misma manera podemos seguir evaluando los intervalos de los atributos latitud y longitud.

```
J48 pruned tree
-----
latitude = '(-inf-21.005122]': '(108.154498-113.516863]' (15898.0)
latitude = '(21.005122-23.755607]'
| id_dispositivo <= 7: '(113.516863-118.879229]' (5505.0)
| id_dispositivo > 7: '(108.154498-113.516863]' (4184.0)
latitude = '(23.755607-26.506092]': '(97.429767-102.792133]' (17404.0/312.0)
latitude = '(26.506092-29.256576]': '(97.429767-102.792133]' (14574.0/1238.0)
latitude = '(29.256576-32.007061]'
| id_dispositivo <= 7: '(118.879229-124.241595]' (51835.0)
| id_dispositivo > 7: '(113.516863-118.879229]' (87052.0/54007.0)
latitude = '(32.007061-34.757545]'
| id_dispositivo <= 6: '(118.879229-124.241595]' (28182.0/7449.0)
| id_dispositivo > 6: '(113.516863-118.879229]' (132392.0/78455.0)
latitude = '(34.757545-37.50803]'
| id_dispositivo <= 2: '(108.154498-113.516863]' (22976.0)
| id_dispositivo > 2: '(113.516863-118.879229]' (196225.0/133636.0)
latitude = '(37.50803-40.258514]': '(113.516863-118.879229]' (2124425.0/58702.0)
latitude = '(40.258514-43.008999]'
| id_dispositivo <= 5: '(118.879229-124.241595]' (4828.0)
| id_dispositivo > 5
| | id_dispositivo <= 8: '(113.516863-118.879229]' (5865.0)
| | id_dispositivo > 8: '(81.34267-86.705036]' (55019.0/31541.0)
latitude = '(43.008999-inf)': '(86.705036-92.067402]' (15148.0/5766.0)
```

Fig. 7. Árbol del algoritmo J48.

MATRIZ DE CONFUSIÓN

La matriz de confusión contiene información acerca de las predicciones realizadas por un método o sistema de clasificación, comparando para el conjunto de individuos en de la tabla de aprendizaje o de testing, la predicción dada versus la clase a la que estos realmente pertenecen[16].

| === Confusion Matrix === | | | | | | | | | | | |
|--------------------------|-------|------|---|-------|---|-------|---------|-------|---|---|-------------------------------|
| | a | b | c | d | e | f | g | h | i | j | <-- classified as |
| 0 | 477 | 0 | 0 | 0 | 0 | 0 | 0 | 1541 | 0 | 0 | a = '(-inf-81.34267]' |
| 0 | 23478 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | b = '(81.34267-86.705036]' |
| 0 | 11365 | 9382 | 0 | 0 | 0 | 0 | 0 | 44903 | 0 | 0 | c = '(86.705036-92.067402]' |
| 0 | 1655 | 196 | 0 | 0 | 0 | 0 | 0 | 63861 | 0 | 0 | d = '(92.067402-97.429767]' |
| 0 | 0 | 0 | 0 | 30428 | 0 | 0 | 0 | 52953 | 0 | 0 | e = '(97.429767-102.792133]' |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 24411 | 0 | 0 | f = '(102.792133-108.154498]' |
| 0 | 0 | 0 | 0 | 1146 | 0 | 43058 | 69567 | 0 | 0 | 0 | g = '(108.154498-113.516863]' |
| 0 | 17463 | 0 | 0 | 404 | 0 | 0 | 2228664 | 7449 | 0 | 0 | h = '(113.516863-118.879229]' |
| 0 | 581 | 0 | 0 | 0 | 0 | 0 | 67564 | 77396 | 0 | 0 | i = '(118.879229-124.241595]' |
| 0 | 0 | 5570 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | j = '(124.241595-inf]' |

Fig. 8. Matriz de Confusión

La confiabilidad de la matriz de confusión, se da cuando la diagonal de la matriz es mayor a los otros valores de su propia columna como lo muestra la línea roja en la figura 8. Aquí nos indica que tenemos un valor en la columna **b** de 23478

registros van estar dentro del rango del atributo longitud (81.34267 – 86705036) y que es el mayor valor que está en la columna **b**, la columna **c** su valor es de 9382 y está dentro del intervalo (86.705036 – 92.067402) y así mismo es el mayor valor que está en esa columna. En esta matriz se está cumpliendo con la condición que su diagonal debe ser mayor.

VISUALIZACION DE RESULTADOS

Una de las etapas del procesado de datos puede ser el mero análisis visual de éstos, en ocasiones de gran utilidad para desvelar relaciones de interés utilizando nuestra capacidad para comprender imágenes. La herramienta de visualización de WEKA permite presentar gráficas 2D que relacionen pares de atributos, con la opción de utilizar además los colores para añadir información de un tercer atributo[14]. Además, tiene incorporada una facilidad interactiva para seleccionar instancias con el ratón.

En la figura 9 nos muestra la visualización de la relación de los atributos longitud y latitud. Las marcaciones de colores nos indican los movimientos que han generado los vehículos en sus diferentes rangos, si nos fijamos bien el color rojo está más remarcado esto indica que en ese rango hay más coordenadas de vehículo. El color azul indica las coordenadas que se encuentran cerca del polo norte, si nos ubicamos en esos puntos aproximados verificamos que la mayoría de coordenadas registradas pertenecen al vehículo 10.

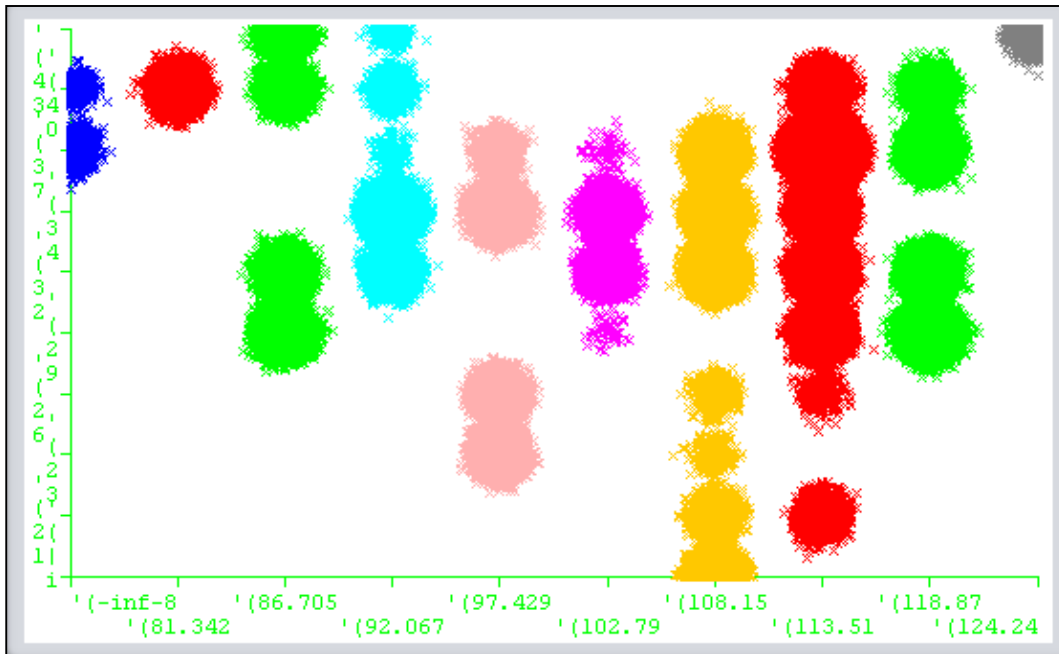


Fig. 9. Visualización de dos variables longitud y latitud

En la figura 10 nos muestra que el atributo longitud va creciendo hacia la derecha y así mismo va aumentando el número de coordenadas por los vehículos.



Fig. 11. Mapa Geográfico[17]

5 CONCLUSIONES

En la presente investigación se analizaron 2781512 registros pertenecientes a 10 vehículos, lo que permitió el análisis profundo de los datos y generar mapas con abundante información que nos permite visualizar los resultados. La utilización de la herramienta Weka ha permitido manipular y generar informes de grandes cantidades de datos, mediante el uso de algoritmos y filtrados de datos. Con estos resultados hemos obtenido identificación de patrones sobre los datos, los cuales se convierten en herramienta fundamental en el análisis de los datos y la realización de proyecciones para el soporte de decisiones. Los retos asumidos acerca de la investigación están vinculados con el análisis y presentación de patrones complejos que deben ofrecer un procesamiento más eficiente, donde el tiempo de respuesta y la visualización de los datos se convierten en el principal objetivo de optimización y desarrollo de algoritmo. Como trabajo futuro proponemos realizar esta investigación con datos geográficos de Ecuador para así determinar cuáles son las rutas más visitadas y que recomendaciones podríamos dar a visitantes sobre los lugares con mayor frecuencia.

REFERENCIAS

- [1] G. Reyes, J. Alvarado, K. Villafuerte, O. Leon, and F. Coral, "Evaluación del marco de trabajo Hadoop y Power View en la Visualización de Trayectorias GPS Vehicular [Evaluation framework Hadoop and Power View display in GPS Vehicle Trajectories]," vol. 16, no. 2, pp. 378–389, 2016.
- [2] J. P. Rigol-Sánchez, M. Chica-Olmo, E. Pardo-Igúzquiza, V. Rodríguez-Galiano, and M. Chica-Rivas, "Análisis e integración de datos espaciales en investigación de recursos geológicos mediante Sistemas de Información Geográfica," *Bol. la Soc. Geol. Mex.*, vol. 63, no. 1, pp. 61–70, 2011.
- [3] D. Description, "User Guide," vol. 3, no. April 2007, 2012.
- [4] V. L. V Lq and G. Pdnlj, "Estado actual de temáticas para el análisis espacial en Resumen Introducción," pp. 59–74, 2014.
- [5] M. X. Dueñas-Reyes, "Minería de datos espaciales en búsqueda de la verdadera información," *Ing. Univ. Bogotá*, vol. 13, no. 1, pp. 137–156, 2009.
- [6] MRA, "Tutorial. Uso basico del API de WEKA," pp. 1–8, 2012.
- [7] B. B. Martínez, "Minería de Datos," 2009.
- [8] M. Rafael, "PostgreSQL," 2010. [Online]. Available: http://www.postgresql.org/es/sobre_postgresql.
- [9] M. Martín, "Manual PostGIS," pp. 1–21, 2007.
- [10] U. Gandhi, "QGis Tutorial and Tips," 2016. [Online]. Available: <http://www.qgistutorials.com/es/docs/introduction.html>.
- [11] M. G. Jiménez and a. Á. Sierra, "Análisis de Datos en WEKA—Pruebas de Selectividad," *It.Uc3M.Es*, pp. 1–9, 2003.
- [12] M. Sevilla-Callejo, "Curso OpenStreetMap," *Osluz.Unizar.Es*.
- [13] M. A. D. E. N. Ormalização, "Manual de Weka," 2012.
- [14] A. De Datos, C. Técnicas, D. E. A. D. E. Datos, and E. N. Weka, "Técnicas de uso de Weka," *Knowl. Creat. Diffus. Util.*, pp. 1–55.
- [15] D. Cangrejo, A. M. Sc, and J. G. A. Ing, "Minería de datos espaciales," *Av. en Sist. e Informática*, vol. 8, no. 3, pp. 71–78, 2011.
- [16] O. Rodriguez, "Aprendizaje Supervisado K - Vecinos más cercanos," 2008.
- [17] Google, "Coordenadas geograficas en Google Maps," 2016. [Online]. Available: <http://www.coordenadas-gps.com/>.