



**Universidad de las Ciencias Informáticas**

**Facultad #6**

*Título: Técnicas de Minería de Datos para el análisis  
de trayectorias de objetos móviles*

*Trabajo de Diploma para optar por el Título de Ingeniero en  
Ciencias Informáticas*

**Autor:** Lisbet Pelayo Grey

**Tutor:** MSc. Yuniel Eliades Proenza Arias

La Habana Junio 2011

## DEDICATORIA

*Dedico esta tesis a las personas que me han atendido durante toda mi vida o parte de ella con amor y comprensión principalmente a:*

*Mis abuelitos por ser mi guía y estar siempre junto a mí en todo momento*

*Mi padres por haberme dado tanto amor y ternura.*

*Mi familia por ser un pilar esencial en mi vida.*

*Mis amigos por apoyarme en mis momentos más difíciles.*

## AGRADECIMIENTOS

*Le agradezco a:*

*Mi abuela por ser la persona más importante en toda mi vida, por darme amor comprensión y apoyo.*

*A mi abuelo por ser la luz que me ilumina el camino a seguir, dedicarme tiempo y ayudarme en todo cuanto yo necesite.*

*A mi madre por ser mi guía, por ser la mejor mamá del mundo, brindarme un amor que sólo en ella puedo encontrar y la ternura que sólo su cuerpo sabe brindar.*

*A mi papá por ser madre y padre al mismo tiempo, por saber comprenderme, tener fe y confiar en mí, gracias por todo papá, representas mucho para mí.*

*A Juanita, mis tías Claudia, Magalis, Clara, Albertina y mis primos Heriam, Oris, Michel por quererme tanto y brindarme amor.*

*A mi familia de Matanzas por ser especial para mí.*

*A mi tutor por haberse comportado como un amigo y ayudarme en todo momento, formándome como una buena profesional.*

*A mi Manolín por ser el mejor hombre que me he encontrado en la vida, eres lo mejor que me pudo haber pasado y a mis amigos Milenis, Pedrito, Pimi, Lisi y Jeem por ser especiales, los admiro y quiero.*

*A mis amigas de la universidad KryKro, Yesuky, Yana, Vivi, Yeni, Noli y Mara por ser pacientes y saber entenderme en todo momento, además de haberme dedicado tiempo y cariño.*

*A mis amigas Yohana, Olamis, Yarelis y Dayexsi porque a pesar de no estar en esta universidad son muy importantes para mí.*

*A mis amistades el aula por hacerme una mejor persona cada día.*

*A Anibal y mis compañeros de baile por haber compartido conmigo tantas horas de ensayo.*

*A Fidel, la revolución y esta universidad por brindarme la oportunidad de estar aquí.*

*A todos los que de una forma u otra me han ayudado en la realización de la tesis.*

## DECLARACIÓN DE AUTORÍA

Declaro que soy el único autor de este trabajo y autorizo al centro GEYSED de la Universidad de las Ciencias Informáticas a hacer uso del mismo en su beneficio.

Para que así conste firmo la presente a los \_\_\_\_ días del mes de \_\_\_\_\_ del año \_\_\_\_\_.

Lisbet Pelayo Grey

Msc. Yuniel Eliades Proenza Arias

\_\_\_\_\_

\_\_\_\_\_

### **DATOS DE CONTACTO**

M.Sc. Yuniel Eliades Proenza Arias

Graduado de Ingeniería Informática en la Universidad de Holguín en el año 2006. Máster en Ciencias en Ingeniería de Software e Inteligencia Artificial por la Universidad de Málaga en el año 2011. Se ha desempeñado como profesor en la Universidad de las Ciencias Informáticas, además de otras tareas asociadas a la producción. Tiene experiencia laboral en la línea del desarrollo de sistemas y la arquitectura de software.

Correo: [yproenza@uci.cu](mailto:yproenza@uci.cu)

## RESUMEN

En la investigación se abordan temas que muestran la importancia de realizar una selección de técnicas de Minería de Datos Espacio-Temporal para su uso en el control de flotas. El principal objetivo del trabajo es identificar diversas técnicas capaces de realizar razonamiento sobre información espacio-temporal de objetos en movimiento aplicables al proceso de control de flotas, estableciendo una analogía entre lo que se denomina flota y objeto móvil.

Las técnicas tratadas en la investigación presentan gran utilidad y funcionamiento para el análisis de trayectorias de objetos móviles, logrando así brindar a través de la extracción de información de bases de datos características esenciales de las entidades en movimiento. La descripción de cada técnica se encuentra enfocado a la funcionalidad que realice, tributando así cada una a diferentes actividades tales como: predicción de movimientos futuros, trabajo con sub-trayectorias similares, agrupamiento, detección de patrones en los datos, entre otras.

Los experimentos evidencian el comportamiento de las técnicas en disímiles escenarios además de, mostrar los resultados arrojados por cada una en circunstancias específicas. Las comparaciones realizadas entre las técnicas resaltan las principales ventajas y desventajas de manera particular, estas se basan en sus características, funcionamiento y resultados, de forma que la selección muestra las técnicas efectivas y precisas a utilizar en el control de flotas.

**Palabras Claves:** Minería de datos espacio-temporal, objetos móviles, control de flotas.

ÍNDICE

**DEDICATORIA..... I**

**AGRADECIMIENTOS ..... II**

**DECLARACIÓN DE AUTORÍA ..... III**

**DATOS DE CONTACTO .....IV**

**TABLA DE FIGURAS.....VIII**

**ÍNDICE DE TABLAS .....IX**

**INTRODUCCIÓN ..... 1**

**CAPÍTULO 1: MARCO TEÓRICO ASOCIADO A LA INVESTIGACIÓN ..... 6**

1.1 Introducción ..... 6

1.2 Conceptos relacionados al dominio del problema. .... 6

1.3 Descripción del proceso de Minería de Datos sobre información espacio – temporal de objetos móviles. .... 11

1.4 Análisis de soluciones existentes..... 13

1.5 Trabajos Relacionados ..... 15

1.6 Conclusiones Parciales..... 16

**CAPÍTULO 2: TÉCNICAS DE MINERÍA DE DATOS ESPACIO-TEMPORAL..... 17**

2.1 Introducción ..... 17

2.2 Metodologías para la detección de regiones de interés (Roi) en una trayectoria..... 17

    2.2.1 Filtración de los lugares candidatos ..... 18

    2.2.2 Puntos de densidad ..... 18

    2.2.3 Agrupación Directa ..... 18

    2.2.4 Método de dos pasos..... 19

2.3 Técnicas de Minería de Datos espacio-temporal para procesar y analizar información.....	20
2.3.1 Técnicas de Minería de Datos espacio-temporal basadas en geometría:.....	21
2.3.2 Técnicas de Minería de Datos espacio-temporal basadas en semántica: .....	29
2.4 Aplicaciones de las técnicas para el control de flotas.....	34
2.5 Conclusiones Parciales.....	34
<b>CAPÍTULO 3: COMPARACIÓN Y SELECCIÓN DE LAS TÉCNICAS .....</b>	<b>35</b>
3.1 Introducción .....	35
3.2 Análisis del funcionamiento de las técnicas. ....	35
Técnica Periódica .....	35
Técnica ObjectGrowth .....	36
Técnica DBSCAN .....	38
Técnica Traclass.....	40
Técnica Traclus .....	44
Técnica DFS_MINE .....	46
3.3 Experimentos realizados.....	48
3.4 Comparación y selección de las técnicas.....	50
3.4.1 Agrupamiento de las técnicas de acuerdo a su funcionamiento .....	50
3.4.2 Comparación y solución.....	51
3.5 Conclusiones Parciales.....	55
<b>CONCLUSIONES GENERALES .....</b>	<b>56</b>
<b>RECOMENDACIONES.....</b>	<b>57</b>
<b>TRABAJOS REFERENCIADOS .....</b>	<b>58</b>
<b>BILBIOGRAFÍA CONSULTADA .....</b>	<b>61</b>
<b>GLOSARIO DE TÉRMINOS .....</b>	<b>63</b>



## TABLA DE FIGURAS

Figura 1: Ejemplo de trayectoria con paradas desconocidas .....	32
Figura 2: Dos trayectorias con paradas desconocidas .....	33
Figura 3: Agrupaciones de objetos en diferentes instantes de tiempo.....	37
Figura 4: Técnica ObjectGrowth.....	38
Figura 5: Densidad accesible .....	39
Figura 6: Agrupaciones y ruidos (DBSCAN).....	40
Figura 7: Agrupamiento basado en regiones.....	41
Figura 8: Agrupamiento basado en trayectorias .....	42
Figura 9: Técnica Traclass .....	43
Figura 10: Comparación de versiones.....	44
Figura 11: Técnica Traclus.....	46
Figura 12: Visualización de la técnica Traclus.....	49
Figura 13: Visualización de la especie Swainsoni .....	49

## ÍNDICE DE TABLAS

Tabla 1: Agrupamiento de las técnicas basado en sus funcionalidades. ....51

## INTRODUCCIÓN

El hombre, en la medida que ha evolucionado, se ha encontrado en la necesidad de almacenar toda la información con la cual trabaja o que recopila para utilizarla posteriormente. La información puede ser procesada y transformada en conocimiento, lo que permite que la misma sea de gran utilidad en diferentes esferas de la vida. Así surgen en 1963 las Bases de Datos (BD) que mantienen los datos almacenados y se puede acceder a ellos de forma directa. Además, los datos pueden ser relacionados entre sí y manipulados por un conjunto de programas con el objetivo de utilizarlos en sistemas de información de una empresa o negocio.

En la actualidad el trabajo que realiza el hombre se ve condicionado en muchos aspectos por la utilización de las Tecnologías de la Informática y las Comunicaciones (TIC), las que proporcionan herramientas que facilitan, en gran medida, su trabajo en distintas actividades tales como: procesamiento de datos, almacenamiento digital de grandes cantidades de información, automatización de tareas y la comunicación entre las personas.

El desarrollo alcanzado se evidencia en el avance científico, técnico y cultural de la sociedad mundial, en el aumento de información en BD y en la posibilidad de que los datos almacenados puedan presentar distintos formatos. Esto ha provocado que la interpretación de la información almacenada necesite de técnicas avanzadas de Minería de Datos (MD). Además, la búsqueda, exploración y procesamiento de los datos cada vez es más complicado de realizar en los medios de almacenamiento. Es por ello que se emplea la MD como soporte de la gestión de datos e información.

La MD es un paso en el proceso de Descubrimiento de Conocimientos en BD (KDD). Además es una forma de buscarle sentido al gran cúmulo de información que puede ser almacenada en la actualidad y se encarga de dos tareas principales: una descriptiva que busca modelos o relaciones existentes en los datos; y otra predictiva que permite identificar nuevos valores basándose en los obtenidos anteriormente.

En Cuba la MD se ha utilizado en diversas empresas con el objetivo de facilitar la exploración en BD, la extracción y procesamiento de información, como ejemplos se pueden destacar el Sistema Cubano de Farmacovigilancia, con amplia utilización en la bioinformática y el análisis de información clínica.

En la facultad 6 de la Universidad de las Ciencias Informáticas se encuentra el centro GEySED que está compuesto por dos departamentos: Señales Digitales y Geoinformática. En el departamento de Geoinformática se desarrollan varios proyectos entre los cuales se puede apreciar Control de Flotas.

Este proyecto posee como misión el desarrollo de Sistemas de Información Geográfica (SIG) para el control de flotas, basados en tecnologías de software libre para su uso en entidades nacionales o extranjeras. Dentro de los objetivos del proyecto se encuentran: el análisis de trayectorias y flujos GPS<sup>1</sup>, el razonamiento sobre datos espacio-temporales, además de la predicción de datos futuros. Hoy día el equipo de desarrollo no cuenta con los conocimientos necesarios para enfrentar los objetivos anteriormente expuestos. El proyecto Control de Flotas presenta los siguientes problemas:

- No se posee una noción acerca de cómo gestionar la información cuando el cúmulo de los datos aumenta grandemente.

La presencia de conocimiento sobre el trabajo con las BD es de gran importancia para cumplir con todas las metas requeridas con el objetivo de desarrollar una aplicación eficiente. Estas son uno de los componentes de gran valor en aplicaciones, la misma está condicionada para aceptar diversos tipos de datos y grandes cantidades de los mismos, por lo que la gestión en ella debe ser profundizada para lograr un buen desarrollo de todo el conocimiento que sea extraído; de ahí la importancia que el equipo de desarrollo posea conocimiento acerca de cómo gestionar la información en las mismas.

- No se tienen elementos que brinden argumentos sólidos acerca de técnicas capaces de razonar el análisis de la trayectoria recorrida por un objeto móvil en el menor tiempo posible.

El equipo de desarrollo debe poseer elementos que le permitan brindar al proyecto la capacidad de utilizar técnicas de MD espacio-temporal y a su vez estas técnicas deben encontrarse capacitadas para razonar la trayectoria recorrida por un objeto móvil en el menor tiempo posible. Todo el trabajo realizado en la investigación y la madurez adquirida en el tema será base de conocimiento para permitir que su incorporación a los SIG tribute al desarrollo de los sistemas para el control de flotas.

- No se conoce cómo predecir movimientos futuros de entidades móviles con la mayor exactitud.

---

<sup>1</sup>**Global Positioning System:** Sistema de posicionamiento global

Actualmente existen técnicas de MD espacio-temporal que se aplican en la información contenida en los grandes almacenes de datos y que, además de mostrar soluciones a diversos problemas, posibilitan el descubrimiento de conocimientos. Hoy día se evidencian diversos trabajos e investigaciones tratando temas importantes sobre las técnicas de MD espacio-temporal, pero en ellos no se hace alusión a la aplicación de las mismas al control de flotas; por lo que se presenta la necesidad de clasificarlas y realizarles una comparación basada en diversos parámetros que responden a las necesidades del proyecto Control de Flotas y lograr así, seleccionar las más idóneas para su utilización.

Teniendo en cuenta lo expuesto anteriormente surge el siguiente **problema de la investigación**: ¿Cómo analizar y procesar información espacio-temporal sobre trayectorias de objetos móviles en el proceso de control de flotas utilizando Sistemas de Información Geográfica?

Para darle solución al problema de la investigación identificado se presenta el siguiente **objetivo general**: Seleccionar técnicas de minería de datos espacio - temporal para su empleo en Sistemas de Información Geográfica para control de flotas desarrollados en la UCI.

Basándose en el problema definido se plantea como **objeto de estudio**: Proceso de minería de datos sobre información espacio-temporal de objetos móviles.

El objetivo trazado concreta el siguiente **campo de acción**: Procesos de análisis de trayectoria y predicción del movimiento de objetos móviles, mediante técnicas de minería de datos, en ambientes restringidos.

La **idea a defender** en la presente investigación es: Si se identifica un conjunto de técnicas de minería de datos para el análisis y procesamiento de información espacio-temporal, clasificadas para su empleo en el desarrollo de aplicaciones en el control de flotas, esto debe permitir su futura incorporación a los Sistemas de Información Geográfica para el control de flotas desarrollados en la universidad.

Con el fin de resolver el problema de la investigación y darle cumplimiento a los objetivos planteados con anterioridad de forma sistemática y ascendente se exponen las siguientes **tareas de la investigación**:

1. Caracterizar las principales técnicas existentes para MD espacio – temporales.

2. Identificar propuestas actuales de técnicas para la MD asociados a objetos móviles.
3. Establecer criterios de clasificación y comparación siguiendo propuestas existentes.
4. Clasificar las técnicas identificadas siguiendo criterios establecidos.
5. Seleccionar las técnicas de relevante impacto sobre el análisis del comportamiento de flotas.
6. Desarrollar un banco de pruebas para las técnicas seleccionadas.

Se esperan como posibles resultados:

- Propuesta de técnicas para la MD espacio-temporales.
- Fundamentos teóricos para la clasificación de técnicas de minería a emplear en aplicaciones para control de flotas.

Para la realización de la investigación se utilizaron los siguientes métodos de la investigación:

### **Métodos empíricos:**

Experimento: Para realizar pruebas que permitan poseer más argumentos que tributen a la clasificación de la(s) técnica(s) más idóneas para el control de flotas.

### **Métodos Teóricos:**

Análisis y síntesis: Para conocer, reflexionar y aumentar los conocimientos acerca de técnicas existentes de MD espacio-temporal para el análisis de trayectorias basándose en la bibliografía consultada, y luego, la síntesis para arribar a conclusiones de la investigación.

Histórico y lógico: Para determinar las tendencias actuales del desarrollo de técnicas de MD espacio-temporal basado en el estudio y evolución de las mismas.

En la presente investigación se desarrolla el tema “Técnicas de Minería de Datos para el análisis de trayectorias de objetos móviles”. La misma está conformada por tres capítulos, los cuales brindan en conjunto la solución al problema de la investigación. En el documento se encuentran definidas técnicas de MD sobre información espacio-temporal que pueden ser utilizadas para el control de flotas. Una vez

definidas y analizadas estas técnicas se plantea cuales de ellas constituyen una estructura idónea y completa a utilizar. La selección de las técnicas está basada en argumentos teóricos y prácticos que son expuestos a medida que avanza la investigación.

## CAPÍTULO 1: MARCO TEÓRICO ASOCIADO A LA INVESTIGACIÓN

### 1.1 Introducción

El presente capítulo contiene conceptos y temas sobre la investigación donde el conocimiento de los mismos es de vital importancia para que el lector logre una mayor comprensión de la investigación. Se abordarán además argumentos que describan el estado actual del objeto de estudio y campo de acción, su seguimiento y utilización en la actualidad. Se presentarán y describirán aplicaciones para el control de flotas existentes a nivel mundial.

### 1.2 Conceptos relacionados al dominio del problema.

Para lograr una correcta asimilación de todo tema a tratar en el documento, se hace necesaria la especificación de diversos conceptos que permiten el cumplimiento de los objetivos y la realización de la investigación. Debido a la creciente información que se maneja en la actualidad, surge la necesidad de darle a la misma un uso racional, correcto y organizado. Este uso proporcionará un mejor aporte a la sociedad y los conocimientos mostrados o adquiridos por las personas tendrán mayor base sobre la cual sustentarse. Con el objetivo de que la información se observe como una herramienta poderosa capaz de transmitir explícitamente argumentos enriquecedores de contenidos importantes en la vida del hombre surge la **gestión de la información**.

*“La gestión de la información es un proceso que incluye operaciones como extracción, manipulación, tratamiento, depuración, conservación, acceso y/o colaboración de la información adquirida por una organización a través de diferentes fuentes y que gestiona el acceso y los derechos de los usuarios sobre la misma” (Curto, 2010).*

Según Bartle la gestión de la información implica (Bartle, 2007):

- Determinar la información que se precisa
- Recoger y analizar la información
- Registrarla y recuperarla cuando sea necesaria
- Utilizarla



➤ Divulgarla

Toda información utilizada debe tener la posibilidad de ser almacenada, con el objetivo de que la misma pueda ser recuperada en caso de pérdida y también evitar su uso por personal no autorizado. Las BD brindan a la humanidad la posibilidad de archivar en memoria cuanta información se desee, ya sea para su utilización momentáneamente o en tiempos futuros.

*“**Base de Datos** se define como un conjunto de información estructurada en registros y almacenada en un soporte electrónico legible desde un ordenador. Cada registro constituye una unidad autónoma de información que puede estar a su vez estructurada en diferentes campos o tipos de datos que se recogen en dicha BD”* (Yunta, 2001).

La información se encuentra ordenada en registros dentro de una BD. Un **registro** *“es un conjunto de campos que contienen los datos que pertenecen a una misma repetición de entidad. Se le asigna automáticamente un número consecutivo (número de registro) que en ocasiones es usado como índice aunque lo normal y práctico es asignarle a cada registro un campo clave para su búsqueda”* (Yunta, 2001).

Las BD ayudan a recuperar la información rápida y flexiblemente, además permiten la ejecución de diversas técnicas estadísticas sobre ellas, las cuales aportan grandes funcionalidades para el trabajo con los datos. Debido al aumento que ha existido en los datos estas técnicas no son las más factibles en la actualidad ya que no dan solución a todos los problemas e incógnitas existentes.

*“Una **BD espacial** es un sistema gestor de BD que maneja datos existentes en un espacio o datos espaciales”* (Lázaro, 2010). Para el trabajo con BD espaciales se tienen en cuenta el espacio, el cual define un patrón de referencia a la hora de determinar su localización y la relación entre los distintos objetos que la componen. El espacio de más utilidad es el espacio físico, ya que este es un patrón manipulable, perceptible y que sirve de referencia. En las BD espaciales se necesitan técnicas capaces de permitir trabajar con grandes cantidades de datos, además de poseer procedimientos para tratar con datos espaciales. Entre los datos que se encuentran inmersos en las BD existen patrones y dependencias entre los mismos.

En la actualidad, la existencia de diversos dispositivos y software que brindan la posibilidad de la captura continua de las posiciones de objetos móviles, posibilita que las BD de objetos en movimiento ocupen en un lugar cimero y hayan acaparado una mayor atención científica y práctica. La representación y manejo de los cambios afines con el movimiento de los objetos son peculiaridades de las BD espacio-temporales, siendo las BD de objetos en movimiento un caso particular de estas.

En las BD de objetos en movimiento los objetos preservan sus identidades, lo que varía en ellos es su ubicación o forma a través del tiempo, también conocido como aspecto geométrico. Esta es una de las diferencias con las aplicaciones espacio-temporales asociadas a los fenómenos geográficos, en las cuales las identidades de los componentes geográficos puede que varíen en el tiempo. En esta rama las aplicaciones que han alcanzado un alto nivel de desarrollo como sistemas de información son aquellas en las cuales los objetos a analizar son automóviles, aviones o cualquier objeto con movimientos regulares.

*“Un **patrón** se refiere a cualquier relación entre los elementos de la BD. Un patrón es interesante en la medida que sea confiable, novedoso y útil respecto al conocimiento y los objetivos del usuario. Además representa conocimiento si su medida de interesante rebasa un cierto umbral, lo cual está basado únicamente en medidas definidas por el usuario” (Aguilera, 2009).*

Entre los patrones espacio-temporales que existen se encuentra una clase denominada “**patrones de movimiento**, que en datos de trayectoria, se refieren a hechos y episodios más destacados expresados por un conjunto de entidades” (Song Li, 2009).

En las BD existe gran cantidad de información oculta, la cual posee un alto grado de argumentos e importancia debido a que brinda el valor real a los datos, además de permitir la toma de decisiones y la comprensión de los fenómenos que rodean a la humanidad. Esta información no puede ser descubierta a través de las conocidas técnicas de recuperación comunes o estándares. Lo que hace posible su revelación es la MD, de ahí que esta técnica de la Inteligencia Artificial juegue un importante papel en la actualidad.

La MD surge como resultado de los grandes adelantos científicos y técnicos desarrollados por la humanidad, permite realizar sobre las BD técnicas que utilizan algoritmos que existen desde años anteriores pero que aún no habían sido implementadas. Estas técnicas brindan respuestas y soluciones a

diversas dificultades que a través del tiempo han nacido a consecuencia de mayores demandas y necesidades expuestas por empresas, además del aumento evidenciado en las tecnologías a utilizar por diversas instituciones.

*“La **Minería de Datos** se encarga de encontrar patrones y relaciones dentro de los datos permitiendo la creación de modelos, es decir, representaciones abstractas de la realidad”* (Vallejos, 2006). Dentro de la MD se definen un conjunto de pasos o áreas de trabajo para lograr una excelencia en la calidad del trabajo a realizar en la misma. Este conjunto de áreas se define como un proceso de integración de las mismas para lograr la identificación de un conocimiento alcanzado a partir de otras BD que tengan un peso en la toma de decisiones.

La MD es una etapa dentro del complejo *proceso de **extracción de conocimiento de BD*** por sus siglas en inglés KDD este *“se encarga de la preparación de los datos y la interpretación de los resultados obtenidos, los cuales dan un significado a estos patrones encontrados. Es el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos”* (Vallejos, 2006).

Como la MD permite el procesamiento de información de forma automática de grandes cantidades de información con el fin de obtener conocimiento útil, la misma brinda la comodidad al usuario del uso de esta información valiosa. Dada la necesidad de crear técnicas eficientes con el objetivo de encontrar conocimiento útil, válido, relevante y nuevo sobre uno o varios fenómenos o actividades es que surge el KDD, siempre bajo la influencia de los crecientes órdenes de magnitud en los datos. Como mismo se le da solución a esta necesidad, existe un profundo interés por mostrar los resultados de dichas técnicas de forma visual o que su representación sea interpretada de una forma bastante clara. Para lograr estas demandas es necesario que la interacción humano-máquina tome un ritmo flexible, dinámico y colaborador, que el resultado de la investigación previa sea interesante y su calidad no sea afectada por mayores volúmenes de datos o por ruido en los mismos y que las técnicas de solución sean lo más robustas posibles. En la presente investigación se estará viendo el desarrollo de esta rama de la Inteligencia Artificial. Se realizará un estudio para aplicar todo lo existente al análisis de trayectorias de objetos móviles.

“La **trayectoria** es la evolución de un objeto que se mueve en el espacio durante un intervalo de tiempo dado” (Stefano Spaccapietra, 2007). Es el trayecto o camino creado por la entidad en movimiento a través del espacio donde se desplaza, la misma no se crea de forma instantánea ya que requiere de un tiempo que se convierte en un aspecto propio de las trayectorias, o sea, que para un intervalo de tiempo correspondiente a una trayectoria definido por  $T_o$  y  $T_n$ , para cualquier instante de tiempo  $T_i$  dentro del intervalo existe una posición en el espacio ocupada por el objeto en movimiento.

Con esto se afirma que toda trayectoria puede ser vista como una función que mezcla instantes de tiempo con momentos en el espacio o como un conjunto de pares (tiempo, ubicación) a causa de que el tiempo es continuo y hay un número infinito de dichos pares en una trayectoria; sin embargo, por cuestiones prácticas deben ser representadas por secuencias finitas de ubicaciones en el espacio referenciadas en un instante de tiempo.

“Un **objeto móvil** es un objeto espacial que se mueve y/o cambia su forma a través del tiempo, además pueden estudiarse su trayectoria o las fuerzas que intervienen sobre él” (Velazco, 2010). Con respecto al tema de los objetos móviles se puede decir que en los mismos se deben realizar aplicaciones para su uso en futuras empresas que necesiten estos argumentos basados en la MD. Existen técnicas de MD que son capaces de brindar a situaciones surgidas, soluciones, por lo que con el transcurso de los años estas han ganado en demanda e importancia para las instituciones.

En la MD existen técnicas que facilitan el trabajo en aplicaciones que poseen este proceso. “Las **técnicas de MD** se aplican sobre información contenida en almacenes de datos. Además se utilizan para descubrir patrones de comportamiento en objetos” (Cerrano Cinca, 2010).

La principal diferencia entre MD y KDD radica que en el descubrimiento de información no existe la necesidad de formular previamente una hipótesis por parte del descubrimiento de conocimientos en BD. Sin embargo la MD está basada en diversas técnicas dentro de las cuales resalta la aplicación automatizada de técnicas que permitan detectar fácilmente patrones en los datos. Esta técnica tiene gran ventaja cuando se intenta explorar datos procedentes de repositorios de gran tamaño y complejidad elevada ya que es mucho más eficiente que el análisis dirigido a la verificación, siendo esta otra de las técnicas existentes. A pesar de las diferencias que existen entre ambas técnicas, estas se encuentran en

continua evolución como resultado de la colaboración entre campos de investigación tales como BD, reconocimiento de patrones, inteligencia artificial, sistemas expertos, estadística, visualización, recuperación de información y computación de altas prestaciones.

En la actualidad se ha agudizado el trabajo vinculado a SIG en los cuales es de mucha utilidad la MD, la cual posibilita un gran avance y ayuda para el buen rendimiento de los sistemas. Un “**Sistema de Información Geográfica** es un conjunto de métodos, herramientas y datos que permiten capturar, almacenar, analizar, transformar y presentar toda la información geográfica y de sus atributos almacenada en una BD espacial” (Gámez, 2009).

### **1.3 Descripción del proceso de Minería de Datos sobre información espacio – temporal de objetos móviles.**

Las MD espacio-temporal utiliza patrones de movimiento capaces no sólo de detectar un período aproximado de movimientos sino que también pueden permitir obtener patrones de movimiento colectivo. Entre los patrones existentes se encuentra el patrón periódico (Periodic), el cual puede ser utilizado para el descubrimiento del comportamiento específico de los objetos en movimiento. Esta le brinda la posibilidad de desarrollar una gran capacidad con respecto a la predicción de movimientos futuros. El patrón Enjambre (Swarm), es capaz de detectar el movimiento en grupo, aunque también se utiliza para predecir comportamientos inusuales de los objetos en movimiento, es decir, si un objeto individual en movimiento abandona temporalmente el grupo siempre y cuando el mismo se encuentre cerca de otros miembros este patrón también será capaz de detectarlo. Existen otros patrones asociados al comportamiento grupal como el Flock (similar al enjambre) y el Follower (seguidor), donde el comportamiento de un objeto en movimiento es similar a otro.

Dentro de la MD se puede evidenciar la **MD espacio-temporal** la cual “se entiende por la búsqueda de patrones de las representaciones concisas de comportamientos interesantes de simples objetos en movimiento o grupos de objetos” (Gámez, 2009). La MD sólo es una etapa dentro del gran proceso de KDD. La autora Vallejos propone las siguientes metas del KDD que inmersas en este proceso constituyen una guía (Vallejos, 2006):

- Procesar automáticamente grandes cantidades de datos crudos.

- Identificar los patrones más significativos y relevantes.
- Presentarlos como conocimiento apropiado para satisfacer las metas del usuario.

Otro punto importante es la **MD de objetos en movimiento** la cual significa la “*detección de periodicidad en el movimiento de objetos y puede ser cíclica (mensual, diaria) o no, lo cual es muy útil en predicción de movimiento*” (Una revisión de Bases de Datos con objetos en movimiento, 2005).

La MD reúne un grupo de técnicas que se encuentran orientadas a descubrir y extraer conocimiento inmerso en una BD. Ésta reúne características fundamentales de ciencias como la Inteligencia Artificial y el análisis estadístico con el objetivo de cumplir sus metas, la misma se encuentra sustentada por estas ramas que brindan argumentos que son capaces de realizar diversas acciones de alta importancia sobre los datos. A través de los patrones adquiridos de la información almacenada se logran resolver problemas que surgen a raíz de situaciones de las cuales no se conoce la solución. Estos problemas resueltos son de predicción, clasificación y segmentación.

Las técnicas de MD son clasificadas de acuerdo a los algoritmos que utilicen. Estas se dividen en dos grandes categorías: supervisadas o predictivas y no supervisadas o de descubrimiento del conocimiento. Cuando se está en presencia de las supervisadas o predictivas se podrá predecir el valor de un atributo de un grupo de datos si se posee conocimiento acerca de los otros atributos conocidos de este grupo. Una vez conocido un atributo de un conjunto de datos se puede realizar un relación entre este atributo y otros que posea el grupo y que ya sean conocidos, esta relación será de ayuda para la predicción de los datos en los que no se conozca el valor del atributo a predecir.

En el desarrollo de una aplicación puede darse el caso de que esta no se encuentre en su estado maduro, por lo que no estará capacitada para brindar soluciones predictivas. En este caso se recurrirá a los métodos del descubrimiento del conocimiento, estos son capaces de descubrir patrones y tendencias en los datos que están siendo utilizados en ese momento, es decir, no requiere de información referente a datos históricos. El descubrimiento de esta información será de gran beneficio y provecho para realizar acciones asentadas sobre una base sólida y de esta forma el beneficio obtenido será brindar un mejor rendimiento en su utilización. De ahí la gran importancia acerca de la necesidad de que todas las aplicaciones posean un amplio y eficiente conocimiento acerca de la extracción de conocimiento en BD.

Para aplicar técnicas de MD, los datos primeramente deben transitar por un proceso nutrido de acciones que permitan anivelar los datos, ya que los mismos pueden provenir de diversas fuentes, lo que posee como consecuencia en diversas ocasiones que no contengan el mismo formato. Una vez preparado los datos ya se pueden obtener resultados que deberán ser evaluados e interpretados, de forma que la utilización de los mismos proporcione una base contundente ante cualquier circunstancia.

El proceso de MD brinda la posibilidad del análisis de trayectorias. Los métodos para la minería de trayectoria poseen muchos escenarios de aplicación, por ejemplo, las técnicas de MD para el análisis de trayectorias no sólo puede detectar automáticamente un período de tiempo aproximado en el que ocurren los movimientos, sino que pueden revelar patrones de movimiento colectivo, además de llevar a cabo la agrupación, clasificación y detección de valores atípicos para el análisis geométrico de las trayectorias. El uso de este análisis aporta grandes ventajas a empresas e instituciones interesadas en explotar esta oportunidad que se abre y da respuesta a disímiles requerimientos actuales.

La población mundial se encuentra rodeada de nuevos avances que traen consigo la necesidad de nuevas estrategias de control, el análisis de trayectorias con uso de la MD es un paso de evolución y contribuye a poseer conocimiento acerca de todo objeto móvil que se traslade. En la Universidad de las Ciencias Informáticas actualmente no se ha desarrollado ningún producto que contenga un eficaz e íntegro manejo de la MD, es decir, que sea capaz de brindar servicios que además de aportar argumentos de conocimientos extraídos de los datos también pueda razonar información extraída de una BD. La ausencia de un módulo capaz de realizar análisis en las aplicaciones conlleva a que no se encuentren capacitadas para valorar información almacenada y arribar a conclusiones. La capacidad de predecir valores en una aplicación es un componente de superioridad y por ende está capacitada para enfrentar adversas situaciones y tener la habilidad de evitar escenarios con problemas que pueden surgir a partir de dificultades existentes.

#### **1.4 Análisis de soluciones existentes.**

Debido a la gran importancia que posee el control de flotas mundialmente surge la necesidad de realizar una investigación que aporte argumentos para la clasificación de diversas técnicas existentes de MD espacio-temporal que puedan ser utilizadas en el control de flotas.

Existen diversos software que se dedican al control de flotas. Estos no poseen ningún razonamiento o análisis capaz de aportar conocimiento extraído de la BD donde se almacena toda la información contenida para su desarrollo. A continuación serán detallados algunos de los sistemas existentes para el control de flotas.

El programa **GlobalSat** (GlobalSat, 2010), apoyado del hardware que lo acompaña permite realizar diferentes acciones para el control de flotas sobre los vehículos que se encuentren registrados. Este sistema posee un mapa que muestra la localización en todo momento de un vehículo, además de especificar cuánto tiempo se ha tomado en cada parada que ha realizado. Esta información que recopila se almacena en la BD para cualquier control o consulta futura que requiera de los mismos argumentos ya obtenidos.

Además del expuesto anteriormente también se puede mencionar a **Micronav** (Micronav, 2010). Este sistema se basa en GPS/GPRS<sup>2</sup> e Internet para proveer un servicio de gestión y orientación de flotas. La utilización del servicio le posibilita disponer de los recursos móviles, mantenerlos controlados y así poder disfrutar de todos los beneficios consecuentes de poseer la información actualizada y poder disponer de ella en todo momento. Este sistema le brinda la asistencia al cliente utilizando la vía de Internet en forma de ASP<sup>3</sup>, el cliente accede al servicio a través de una aplicación determinada o por el Explorador de Windows por Internet. Micronav brinda un conjunto de ventajas que son utilizadas en el beneficio de la sociedad y que no se encuentran presentes en otros sistemas que realizan o poseen el mismo objetivo.

Una propuesta de SIG a mencionar es el sistema **MovilWeb** (Lic. Osmani Herrera González, 2008). Es importante recalcar de esta herramienta que basa su funcionamiento sobre Internet Information Server (IIS). Se utiliza para el control de flotas y fue desarrollado en Cuba, poniéndola en práctica desde el año 2006. La herramienta permite un conjunto de acciones que tributan al avance científico del país, entre ellas se encuentran:

- Seguimiento de móviles
- Análisis de la ruta y trayectoria histórica del movimiento de un vehículo

---

<sup>2</sup> **General Packet Radio Service:** Paquete General de Servicios de Radio

<sup>3</sup> **Application Service Provider:** Proveedor de servicios de aplicaciones



- Análisis de las distancias recorridas por un objeto móvil
- Visualización de la trayectoria de los vehículos

Esta aplicación posee en su BD más de 7 000 vehículos de distintas bases de transporte del centro de Cuba. A pesar de que MovilWeb posee en su estructura una implementación de componentes para utilizar los servicios Web que aporta la Infraestructura de Datos Espaciales de la República de Cuba (IDERC) y utiliza como Gestor de BD al SQL 2000, se ha comenzado a desarrollar una versión más sencilla, la que permitirá más acciones de las que ya realiza MovilWeb como realizar el seguimiento de vehículos en tiempo real.

Entre los diversos sistemas vistos se puede mencionar el sistema **ArcLogistics** (Tirado, 2010), el cual constituye una solución integral para el control y administración de flotas. Se basa en información almacenada acerca de los vehículos como: las calles, tráfico y horario, para lograr de esta forma una gestión de las rutas de manera inteligente y mejorando las posibilidades, ya que se apoya en herramientas informáticas que aportan argumentos capaces de dar solución a los problemas de planificación, minimización de costos y maximización de la productividad.

Todas las soluciones mencionadas con anterioridad poseen provechosas ventajas. Una característica en común para todos es que no se encuentran basados ni fundamentados sobre ninguna técnica de MD, por lo que la utilización de este proceso en ellas es nula. Esta sería de gran ayuda con el objetivo de ahorrar recursos en determinadas acciones que son realizadas por personal, debido a la ausencia de técnicas capaces de razonar y analizar los datos almacenados.

## 1.5 Trabajos Relacionados

Actualmente no se conoce ninguna investigación que posea una selección y clasificación de técnicas de MD espacio-temporal para el control de flotas. De aquí la importancia del desarrollo del presente trabajo de investigación, el cual aportará argumentos para la clasificación de las diferentes técnicas de MD para el uso en control de flotas, logrando así integrar las mismas a un SIG capaz de alcanzar todas las metas no obtenidas por los software existentes a causa de la ausencia de razonamiento en los mismos. También dará solución a las disímiles incógnitas surgidas como consecuencia de la demanda por parte de la población de aplicaciones que integren el manejo de información espacial, geográfica y temporal acerca

de los objetos móviles. Además, se incorporarán nuevos elementos al cuerpo de conocimiento de la MD espacio-temporal y al control de flotas al desarrollar elementos teóricos referentes al tema.

## **1.6 Conclusiones Parciales**

En el presente capítulo se han expuesto diversos conceptos y elementos teóricos que sustentan el problema de la investigación a tratar, asociados a disímiles técnicas de MD de información espacio-temporal para el uso en control de flotas, facilitando argumentos que permitan comprender mejor los términos a manejar en la investigación. La descripción realizada del objeto de estudio y todo lo asociado al dominio del problema ha posibilitado la ardua interpretación de elementos necesarios para el cumplimiento de los objetivos de la investigación. El estudio de aplicaciones existentes para el control de flotas, evidencia la necesidad del desarrollo del trabajo brindándole al lector una visión de los aspectos teóricos manejados además de, asegurar los elementos necesarios para el posterior desarrollo de la investigación.

# CAPÍTULO 2: TÉCNICAS DE MINERÍA DE DATOS ESPACIO-TEMPORAL

## 2.1 Introducción

El presente capítulo contendrá información acerca de técnicas utilizadas para la extracción de conocimiento de datos contenidos en grandes almacenes. Se realizará una descripción de las mismas resaltando su función en el análisis de trayectorias de objetos móviles, además de mostrar características capaz de identificarlas en cuanto a cualidades que posean una vez insertadas en las aplicaciones para el control de flotas. Toda la argumentación brindada será la base para la comprensión del funcionamiento de estas diferentes técnicas.

## 2.2 Metodologías para la detección de regiones de interés (Roi) en una trayectoria.

Las trayectorias deben ser particionadas antes de su posterior análisis debido a varias razones; la primera y más importante es poder trabajar con sub-trayectorias de los objetos en movimientos que se correspondan con el sentido semántico del viaje completo en lugar de sólo una serie de coordenadas. Por ejemplo, un pájaro que se ha apartado de la migración hará una parada en alguna parte por algún tiempo para alimentarse, otra para descansar y así continuará hasta culminar su viaje. Luego de haber particionado la trayectoria esta puede ser dividida en Roi.

En segundo lugar, una trayectoria puede estar compuesta por sub-trayectorias similares. Por ejemplo, una bandada de aves migratorias pueden dividirse en dos rebaños después de compartir un largo viaje juntas para dirigirse a dos hábitats diferentes. Por último, debido a la innata incertidumbre, la trayectoria debe dividirse cuando se requiera la mayor precisión posible en el análisis. Por ejemplo, dos puntos de observación consecutivos que anteriormente eran iguales pasado un día poseen gran diferencia, mientras que en otros con unos pocos segundos transcurridos se observa igual o mayor diferencia.

Los Roi pueden ser convenientemente seleccionados de una BD de los lugares candidatos o de un sub-conjunto que cumpla ciertos criterios determinados. Sin embargo, en algunos casos no se tiene esta información de antemano, por lo tanto, tiene que ser derivada de alguna manera. Existen diferentes metodologías que incluyen técnicas de MD espacio-temporal para la detección de un Roi.

### **2.2.1 Filtración de los lugares candidatos**

La técnica SMOT (Paradas y movimientos de las trayectorias) está basada en la idea de encontrar paradas en lugares candidatos. Una parada se puede encontrar si el objeto que se mueve se queda el tiempo suficiente en una característica geográfica (lugar importante) estudiada y de esta forma podrá ser registrada. Más adelante en el epígrafe 2.3 se presentará una caracterización de esta técnica con un mayor nivel de detalle para la comprensión de la misma.

### **2.2.2 Puntos de densidad**

Los Roi pueden ser extraídos de un mapa de densidad de puntos, donde la densidad alta naturalmente corresponde a una alta importancia. Sin embargo, este método posee varios inconvenientes; en primer lugar, el valor umbral de la importancia es arbitrario, en segundo lugar, por Roi diferentes, la densidad umbral debe ser diferente, esto se evidencia cuando por ejemplo, la ubicación del banco de la ciudad tiene mayor densidad que la empresa donde está trabajando una persona, y la ubicación de la empresa puede tener mayor densidad que la de la casa de esa persona. Sin embargo, si la densidad umbral se fija de acuerdo a la casa, disímiles lugares como las carreteras que pasan por su alrededor serían incluidas sin tener ningún significado semántico en la mayoría de los estudios. Esta desventaja es especialmente evidente en el seguimiento de animales.

### **2.2.3 Agrupación Directa**

Para aplicar esta metodología se utiliza el método que consiste en descomponer toda la trayectoria  $P$  en  $n$  sub-trayectorias de tipo  $P$ , a continuación, todos los lugares de la sub-trayectoria se agruparán en un grupo  $G_w$ .  $P$  es dependiente de los datos y no tiene ningún valor definido, esto se puede evidenciar en el siguiente ejemplo:  $P$  es igual a "un día" en una aplicación de control de tráfico ya que muchos vehículos tienen patrones diarios, mientras que en los comportamientos de migración anuales de los animales  $P$  puede tomar el valor  $P$  es igual a "un año". La técnica espacio-temporal DBSCAN se aplica para realizar una búsqueda en el grupo  $G_w$ . Además, para evitar generar demasiados grupos grandes, los grupos que son más grandes que el umbral especificado automáticamente se dividen en dos. Este método difiere de

los demás debido a que se basa en las características dinámicas de los Roi.

### **2.2.4 Método de dos pasos**

En la mayoría de los casos, los Roi se encuentran a menudo en dos pasos: el primero es descubrir los lugares de mayor densidad, es decir los más importantes. Luego de ser detectados los puntos en el espacio, se extraen para que los representen de manera sucinta en un Roi.

#### **2.2.4.1 Lugares importantes**

Diferentes conjuntos de datos proporcionan varias oportunidades para encontrar lugares significativos. Para el seguimiento de las personas, es probable que, al menos para la mayoría de la gente en la ciudad, los lugares que podrían considerarse significativos serán dentro de los edificios donde las señales GPS no llegan. Esto significa que habrá un flujo de datos registrados hasta que el usuario entra en un edificio, luego un intervalo de tiempo, y luego la reanudación de los datos cuando el usuario sale del edificio. Se define como un lugar importante al registro de un intervalo de tiempo  $t$  entre un punto y el punto anterior detectado por un GPS. Si un punto es visitado por un objeto varias veces, una vez que se cuenta, se dice que este lugar importante que se encuentra, es significativo para el público en lugar de decir que es significativo para objetos individuales.

Para reducir la complejidad computacional, con el objetivo de resolver el problema antes planteado, se propuso una técnica llamada CB-SMOT (Agrupaciones basadas en paradas y movimientos de trayectorias) para encontrar lugares interesantes. Este método se basa en que las partes de una trayectoria en la que la velocidad es menor que en otras partes de la misma trayectoria corresponden a lugares de interés. Esto se puede evidenciar en una aplicación de turismo, donde la trayectoria de un turista sería algo así como: visita a un importante monumento, visitar un museo, ir a su hotel, ir a un club nocturno y regresar al hotel. Es probable que su trayectoria tenga una velocidad más baja en torno a estos lugares con respecto a otras partes de la trayectoria en la que se movía de un lugar a otro. Luego la técnica espacio-temporal de agrupamiento DBSCAN basado en la densidad de los lugares importantes se adapta a los datos de la trayectoria. Estas técnicas serán vistas posteriormente en el epígrafe 2.3.

### 2.2.4.2 Roi

Debido a que las mediciones de múltiples ubicaciones tomadas en la misma ubicación física pueden variar, el registrador deberá asegurarse de no tomar las mismas coordenadas incluso si el usuario se detiene en ella durante diez minutos, precisamente en el mismo punto todos los días. Los lugares importantes representan información muy fina que es difícil de manejar adecuadamente debido a su gran número; por esto se han creado grupos de lugares utilizando una técnica de agrupamiento. Los grupos resultantes son nombrados regiones para hacer una diferencia del término lugar.

### 2.3 Técnicas de Minería de Datos espacio-temporal para procesar y analizar información.

A continuación se describen métodos que clasifican las técnicas de MD espacio-temporal, agrupándolas así en dos grupos.

Métodos de MD espacio-temporal (Bogorny, 2010):

Dos enfoques:

- Basada en la geometría de MD espacio-temporal :
  - Métodos de agrupamiento basado en la densidad
  - Concentración en la analogía física
  - Considerando sólo las propiedades geométricas de las trayectorias (el espacio y el tiempo)
- Basados en la semántica de MD espacio-temporal
  - Los patrones se calculan en base a la semántica de los datos.
  - Las trayectorias son pre-procesadas para enriquecer los datos.

Los dispositivos móviles han tenido un gran auge en los últimos años, ya que estos generan grandes cantidades de datos de trayectorias. Las trayectorias de los coches, los seres humanos, aves u otros objetos son un nuevo tipo de datos que pueden ser muy útiles en la toma de decisiones en ámbitos de diferentes aplicaciones. Estos datos, sin embargo, tienen muy poca o ninguna semántica de los puntos de muestreo, por lo que la extracción de análisis y el conocimiento de los puntos de muestreo en la trayectoria son muy difíciles desde el punto de vista del usuario, presentándose esta problemática y la

existente necesidad emergente de nuevos modelos de datos, técnicas de manipulación y herramientas para extraer el significado de patrones a partir de estos datos.

### **2.3.1 Técnicas de Minería de Datos espacio-temporal basadas en geometría:**

Las técnicas que a continuación se describen presentan características que evidencian presencia de razonamiento y análisis de los datos. Con su utilización se puede realizar acciones como la prevención de eventos futuros, la comprensión de datos, entre otras que demuestran su utilidad en la sociedad.

#### **Técnica Periódica**

Los comportamientos periódicos pueden proporcionar una explicación profunda y concisa sobre el movimiento de una entidad móvil. Por ejemplo, los movimientos de animales se podrían resumir con varios comportamientos periódicos diarios y anuales. Estos comportamientos periódicos también son útiles para la comprensión de datos de movimiento, puesto que son el resumen de los movimientos que se pueden utilizar para reemplazar los datos originales para ahorrar espacio. Por otra parte, los comportamientos periódicos son muy utilizados en la predicción de movimientos futuros, especialmente para un tiempo prolongado. Al mismo tiempo, si un objeto no puede seguir el comportamiento periódico que regularmente lleva, esto podría ser una señal de un cambio anormal o un accidente en él.

Sin embargo, la minería de comportamientos periódicos a partir de los datos de un objeto en movimiento es un problema difícil. En ocasiones es casi imposible extraer los comportamientos periódicos de los datos en bruto provocando así que esto se convierta en una tarea muy complicada. La minería de comportamientos periódicos puede reducir la brecha entre los datos brutos y la comprensión semántica de los datos.

Existen dos sub-tareas en la minería de comportamientos periódicos, la detección de los períodos y la minería de los comportamientos periódicos. Se propone una técnica de dos etapas denominada Periódica, donde el procedimiento general de la técnica se desarrolla en dos partes y cada etapa cubre los objetivos de una sub-tarea. En el primer algoritmo (Zhenhui Li, 2007) se detectan los períodos para cada punto de referencia.

La primera etapa, se enfoca en la detección de todos los períodos en el movimiento. Teniendo en cuenta los datos en bruto, se utiliza el método del kernel (Zhenhui Li, 2007) para descubrir los lugares de referencia, es decir, puntos de referencia. Se utiliza el método de Fourier transformado (Zhenhui Li, 2007) para la detección de los períodos. Sin embargo, puede que no exista ninguna señal clara que corresponda al período correcto, porque este método es sensible al ruido espacial. Si el objeto no sigue más o menos la misma ruta todos los días, el período apenas puede ser detectado. Sin embargo, en casos reales, algunos objetos repiten exactamente la misma ruta en los movimientos periódicos.

En caso de que se tenga noción de los puntos de referencia se presentan varias ventajas. En primer lugar, se filtra el ruido espacial que son puntos que no pertenecen a ninguna agrupación, y por lo tanto no forman parte de la información, por lo que se pueden presentar como interferencia. Al resolver este inconveniente se convierte el problema de la detección de períodos de un espacio de dos dimensiones (espacial) a un espacio de una dimensión (binarios) para cada punto de referencia, donde al menos uno se asocia con cada período; es decir, mientras se siga un patrón regular hay un período asociado con el estudio. Todos los períodos en el movimiento pueden ser detectados si se trata de descubrir los períodos en cada punto de referencia.

En segundo lugar, se puede detectar varios períodos en el movimiento. Considerando la situación que existe un período diario con un punto de referencia y un período semanal con otro punto de referencia, es posible que sólo el período "día" sea descubierto por el período más corto que se repita más veces, pero si vemos el movimiento de dos puntos de referencia por separado, los dos períodos pueden ser detectados de forma individual. En tercer lugar, si cada comportamiento periódico se asocia a algunos lugares de referencia, todos los períodos se pueden encontrar a través de puntos de referencia.

En la segunda etapa se tendrán en cuenta los puntos de referencia con el mismo período a la vez con el fin de obtener comportamientos periódicos más concisos e informativos. Debido a que un comportamiento sólo puede existir en un movimiento parcial, puede haber varios comportamientos periódicos con el mismo período.

Si existen  $K$  comportamientos periódicos subyacentes, hay muchas maneras de agrupar los segmentos en grupos  $K$  con la medida de distancia definida. Sin embargo, el número de comportamientos periódicos



subyacentes ( $K$ ) es generalmente desconocido. Por lo tanto se propone un método de agrupación jerárquica para el grupo de los segmentos, mientras que al mismo tiempo, se debe determinar el número óptimo de comportamientos periódicos. En cada iteración del agrupamiento jerárquico dos grupos con la distancia mínima se fusionan. Cuando dos grupos se fusionan, el nuevo clúster hereda los segmentos que pertenecen al original.

### **Técnica ObjectGrowth**

El estudio y análisis de datos útiles en movimiento se puede enfocar a la mudanza de varias agrupaciones de objetos, con el objetivo de encontrar un grupo de objetos en movimiento que viajan juntos esporádicamente. El descubrimiento de estas agrupaciones ha facilitado el estudio a fondo del comportamiento de los animales, la planificación de rutas y el control de los vehículos. Un grupo de objetos en movimiento se puede definir en dos dimensiones espaciales y temporales: los objetos en movimiento deben estar geoméricamente cerca el uno del otro y deben permanecer juntos por lo menos en algún tiempo de mínima duración. Esta limitación del tiempo puede provocar la pérdida de interesantes grupos de objetos en movimiento.

El tiempo es de gran influencia y muy importante para darse cuenta cuando los objetos viajan juntos. En ocasiones algunas agrupaciones de objetos en movimiento que comparten la misma trayectoria geoméricamente no viajan juntas en todo momento, sin embargo, en diferentes marcas de tiempo se reúnen para viajar en agrupación.

Existe un patrón de movimiento llamado enjambre visto en el capítulo anterior en el epígrafe 1.3, este es una generalización para mover grupos de objetos. Precisamente enjambre es un grupo de objetos en movimiento que contiene algunos individuos que están en el mismo grupo al menos durante varios instantes de tiempo. Si se denota este grupo de objetos en movimiento como  $O$  y el conjunto de estas marcas de tiempo como  $T$ , un enjambre es un par  $(O, T)$  que satisface las restricciones anteriores.

El descubrimiento de enjambres plantea un nuevo problema que debe ser resuelto por medio de técnicas diseñadas específicamente. Este hallazgo genera un enorme espacio de búsquedas, por lo que se utiliza para su detección un eficiente método llamado ObjectGrowth. Además de la regla Apriori poda (Zhenhui Li, 2009) que se utiliza comúnmente, presenta una nueva con versiones anteriores de la regla de poda;

esta utiliza un paso de comprobación simple para detener la búsqueda en caso de que no se necesite buscar más y la misma podría abarcar varios casos redundantes al mismo tiempo. Este método también trabaja con enjambres cerrados, los cuales son los objetos que se mantienen en movimiento juntos durante varios instantes de tiempo.

Estas reglas de poda cortan una gran porción de los candidatos, sin embargo el número restante de enjambres cerrados puede ser grande. Para evitar consumir tiempo en la verificación de los pares cerrados se realiza la comprobación de cierre sobre la marcha; poniendo en práctica esta regla se reduce la necesidad de utilizar espacios para el almacenamiento de los candidatos y también se evita consumir tiempo extra realizando la verificación de cierre.

ObjectGrowth se desarrolla para la extracción eficiente de enjambres cerrados. Dos reglas de poda se desarrollan para reducir eficazmente el espacio de búsqueda y un control de paso de cierre está integrado en el proceso de búsqueda de salida de los enjambres cerrados de manera inmediata.

La idea básica de la técnica es, encontrar enjambres cerrados extraídos de los sub-conjuntos de los objetos en movimiento, a pesar de que el espacio de búsqueda sigue siendo enorme para enumerar los conjuntos de objetos (ObjectSet). La primera regla se utiliza para detener la trayectoria de algunos sub-árboles cuando se está en presencia de más de un recorrido que no satisface las restricciones del instante de tiempo. La segunda es hacer uso de la propiedad de cierre donde se comprueba para ver si hay un súper-conjunto de los ObjectSet actuales que tengan el mismo ajuste del instante de tiempo correspondiente al actual. Anteriormente se utilizaban más de una regla de poda, pero solamente con estas dos reglas, el tamaño del espacio de búsqueda se puede reducir significativamente. Este método presenta una gran eficacia y eficiencia demostrada con datos reales en prueba.

### **Técnica DBSCAN**

La técnica de agrupamiento DBSCAN (Density-Based Spatial Clustering of Applications with Noise) es un algoritmo de agrupamiento espacial de datos con ruido basado en la densidad. DBSCAN sólo requiere de dos parámetros de entrada y apoya al usuario para que pueda determinar un valor adecuado para estos.

Para que esta técnica funcione adecuadamente, habría que conocer los parámetros adecuados *Eps* (distancia mínima entre los puntos) y *MinPts* (cantidad mínima de puntos que debe tener un vecindario) (Martin Ester, 1996) de cada grupo y por lo menos un punto de la agrupación respectiva. De esta forma se pueden recuperar todos los puntos que sean de densidad alcanzable, lo que significa que los puntos pertenecen al vecindario *Eps* y que la cantidad de puntos del vecindario sea mayor o igual que *MinPts*. DBSCAN utiliza los valores globales de *Eps* y *MinPts*, es decir, los mismos valores para todas las categorías. Los parámetros de densidad de los grupos menos poblados son buenos candidatos para estos valores de los parámetros globales especificando la densidad más baja que no se considera ruido.

Para encontrar un grupo, DBSCAN comienza con un punto  $p$  arbitrario y recupera todos los puntos de densidad alcanzable. Si  $p$  es un punto central, este procedimiento obtiene una agrupación. Si  $p$  es un punto de la frontera, no tienen los puntos una densidad alcanzable desde  $p$  por lo que la técnica visita el siguiente punto de la BD.

Utilizando los valores globales de *Eps* y *MinPts* DBSCAN puede combinar dos grupos en una agrupación, si estos dos grupos de diferente densidad están cerca el uno del otro. Dos conjuntos de puntos que tienen al menos la densidad del conjunto menos poblado serán separados unos del otro sólo si la distancia entre los dos grupos es mayor que el parámetro *Eps*. En DBSCAN puede ser necesaria una llamada recursiva para detectar agrupaciones con valores mayores que *MinPts*. Sin embargo esta llamada no constituye una desventaja en la técnica ya que la misma se realiza de manera práctica y eficiente.

### **Técnica Traclclass**

Se dice que una región en un espacio de dos dimensiones es homogénea si y sólo si en una clase *ClaseA* la trayectoria se encuentra particionada por  $N$  trayectorias y mayores que  $N$  trayectorias dentro de la región pero las demás clases no. La clase *ClaseA* se le denomina clase principal de la región y a las demás clases se le denomina clases sin importancia. El valor  $N$  es la población mínima de la clase principal en una región homogénea. Las clases son especificadas por el usuario según el tipo de trayectoria.

Una agrupación basada en regiones es un conjunto de trayectorias particionadas de la clase principal en una región rectangular homogénea. Un grupo basado en regiones tiene muchas trayectorias particionadas de la clase principal, pero muy pocas trayectorias particionadas de otras clases menos importantes.

Un grupo basado en trayectoria debe crecer utilizando particiones de la trayectoria de la misma clase. El interés real es garantizar que un grupo se derive de una sola clase. Es obvio que las trayectorias comunes sub-compuestas por diferentes clases no son de utilidad para la clasificación.

Antes de la agrupación, cada trayectoria se divide en un conjunto de trayectorias particionadas. En primer lugar, la agrupación basada en regiones se realiza de forma recursiva, siempre y cuando se encuentran regiones homogéneas de tamaño razonable. Si las trayectorias particionadas no están cubiertas por regiones homogéneas se pasa al siguiente paso. En segundo lugar, el agrupamiento basado en trayectoria se lleva a cabo siempre y cuando las agrupaciones encontradas constituyan agrupaciones que son poco utilizadas o menos útiles.

Traclus recursivamente cuantifica las regiones que no son homogéneas. Esta cuantificación recursiva permite encontrar más grupos basados en regiones. Si las regiones contienen muy pocas o ninguna trayectoria particionada, de ellas no se investiga más, ya que no proporcionan información significativa para la clasificación. Después de que son descubiertas todas las posibles regiones homogéneas, las regiones adyacentes se pueden combinar para formar regiones más grandes. La técnica más bien se centra en clasificar un grupo de sub-trayectorias en lugar de clasificar las trayectorias completas.

### **Técnica Traclus**

Existen técnicas que agrupan trayectorias similares de un grupo en su conjunto, por lo tanto estas técnicas descubren trayectorias comunes. Descubrir sub-trayectorias comunes es muy útil en muchas aplicaciones, sobre todo si se tienen regiones de especial interés para el análisis. La técnica divide una trayectoria en un conjunto de segmentos de línea y, a continuación, los grupos de segmentos de línea similares forman un clúster. La principal ventaja de esta estructura es descubrir sub-trayectorias comunes de la trayectoria. Basándose en esta estructura de partición, se desarrolla una técnica (Jae-Gil Lee, 2007) de agrupación de trayectoria llamada Traclus. La técnica consiste en dos fases: separación y agrupación. Para la primera fase, se presenta un algoritmo de partición de trayectoria utilizando el principio de longitud

mínima de descripción (MDL) y para la segunda fase, se presenta un algoritmo de agrupación (Jae-Gil Lee, 2007) de la densidad basada esta en segmentos de línea. Los resultados experimentales demuestran que Traclus descubre correctamente sub-trayectorias comunes de los datos de la trayectoria real. Esta técnica realiza agrupaciones de trayectorias en su conjunto, es decir, la unidad básica de la agrupación es toda la trayectoria.

### **Técnica DFS\_MINE**

El problema de la minería en los patrones espacio-temporales es encontrar secuencias de acontecimientos que se producen frecuentemente en conjuntos de datos espacio-temporales. Los conjuntos de datos espacio-temporales almacenan la evolución de los objetos en el tiempo. Ejemplo de esto pueden ser las secuencias de imágenes de un sensor de una región geográfica, los datos que describen la ubicación y el movimiento de objetos individuales con el tiempo, o datos que describen la evolución de los fenómenos naturales, como la cobertura forestal.

Los patrones descubiertos son secuencias de eventos que se producen con mayor frecuencia. Se presenta DFS\_MINE como una nueva técnica (Ilias Tsoukatos, 2001) para una rápida extracción de patrones frecuentes espacio-temporales y probada sobre datos. DFS\_MINE, como su nombre indica, utiliza un enfoque de la búsqueda en profundidad que permite el rápido descubrimiento de prolongados patrones secuenciales. Esta técnica realiza exploraciones en la BD para descubrir secuencias en lugar de confiar en la información almacenada en la memoria principal, que tiene la ventaja de que la cantidad de espacio requerido es mínimo.

En enfoques anteriores utilizan como método la búsqueda del primero en profundidad, pero no es eficiente para el descubrimiento de largas secuencias frecuentes. Por otra parte, se requiere almacenar en la memoria principal todas las ocurrencias de cada secuencia en la BD y, como resultado, la cantidad de espacio que se necesita es bastante grande. Los experimentos demuestran que los costos de I/O (entrada y salida) en los análisis que se realizan sobre la BD se ven compensados por la eficiencia del enfoque DFS-Deep First Search (Búsqueda en Profundidad), ya que es el que asegura que el descubrimiento de largos patrones frecuentes sea rápido.

Las características principales de este enfoque son:

- DFS\_MINE utiliza la teoría de la red (lattice) (Ilias Tsoukatos, 2001) como método para descomponer el espacio de búsqueda original.
- Sigue el concepto de búsqueda del primero en profundidad, es decir, se trata de descubrir frecuentes secuencias de longitud  $k$  sin agotar todas las secuencias frecuentes de longitud  $k-1$ . Se utiliza la información sobre secuencias frecuentes ya descubiertas para extraer secuencias de mayor longitud. Se retrocede al igual que en DFS, a las secuencias de menor longitud cuando todas las secuencias de la ruta de acceso del camino de red elegido llegan a ser no frecuente, y garantizar así el rápido descubrimiento de largos patrones frecuentes.
- DFS\_MINE no enumera todas las secuencias frecuentes en la BD y descubre rápidamente sólo las máximas secuencias frecuentes.
- No tiene como objetivo reducir al mínimo el análisis en las BD. Realiza análisis en la BD para determinar la frecuencia de un conjunto de secuencias. A pesar de ello, se consigue una rápida detección de las secuencias frecuentes gracias a la utilización de la estrategia de DFS.
- No requiere de enormes cantidades de memoria, sólo se necesita un espacio mínimo para almacenar dos estructuras (la lista de un máximo de secuencias frecuentes (MaxFreqList) y la lista de un mínimo de secuencias no frecuentes (MinNonFreqList)) que permiten la representación eficiente del espacio de búsqueda. Los experimentos demuestran que DFS\_MINE supera a todas las soluciones existentes en cuanto a tiempo y espacio se refiere, sobre todo cuando las secuencias a ser descubiertas son bastante largas.

Teniendo en cuenta el umbral mínimo de apoyo ( $\text{min\_sup}$ ), explora la BD una vez en busca de ocurrencias de elementos. Puede que los elementos frecuentes se mantengan en la lista de elementos frecuentes (FreqItems) o que los elementos se inserten en FreqItems en un orden específico, no necesariamente alfabético. Cada vez que se vayan a analizar y utilizar los elementos frecuentes ubicados en FreqItems en pasos posteriores, se respeta el orden que tengan los elementos. Para generar todos los candidatos de dos secuencias, se cruzan todos los elementos frecuentes entre sí en todas las combinaciones posibles, este conjunto es también examinado contra la BD. Las secuencias frecuentes se insertan en la lista de máxima secuencias frecuentes (MaxFreqList) y las secuencias no frecuentes son insertadas en la lista de mínimo de secuencias no frecuentes (MinNonFreqList). La técnica utiliza las

secuencias frecuentes para extraer los patrones frecuentes. Las secuencias no frecuentes se utilizan para la poda de las secuencias más largas no frecuentes.

La estrategia DFS\_MINE es ideal para la minería en varios niveles, porque utiliza los resultados del nivel anterior con el fin de descubrir rápidamente las secuencias frecuentes del siguiente nivel. La idea principal es que si se descubre rápidamente una secuencia de secuencia  $k$ , entonces no es necesario perder el tiempo examinando todas sus sub-secuencias, porque se está seguro de que son frecuentes. Una  $k$ -secuencia frecuente se cruza con todos los elementos frecuentes para generar todos los candidatos  $(k+1)$ -secuencias, que luego se analizan contra la BD. La única información que DFS\_MINE almacena para cada secuencia es el "conjunto no utilizado" de las mismas.

### 2.3.2 Técnicas de Minería de Datos espacio-temporal basadas en semántica:

Las técnicas SMOT y CB-SMOT se encuentran vinculadas al sentido y tiempo de las trayectorias. Permiten analizar recorridos realizados y extraer información valiosa para el usuario a partir de estos. Trabajan sobre los datos de las trayectorias y con información obtenida por el usuario, lo que en algunas ocasiones se convierte en una desventaja para estas técnicas.

#### Técnica SMOT

Actualmente hay una gran y creciente necesidad de establecer mecanismos para el análisis eficiente y conocimiento profundo de la extracción de datos. Esto se debe al gran auge que ha ido tomando el almacenamiento de los datos de objetos en movimiento y por ende la presente necesidad de su utilización en la sociedad.

Las funciones más generales para obtener trayectorias semánticas se llaman *generaS* y *generaM*, que, respectivamente, calculan las paradas y los movimientos de las trayectorias. A través de estas funciones, el usuario puede seleccionar un sub-conjunto de datos que le interese y proporcionar la información geográfica que es relevante para el proceso de descubrimiento. Ambas operaciones tienen como entrada los siguientes parámetros:

- El *método* para calcular las paradas y los movimientos.
- Las *paradas candidatas*.
- La *zona de tope* alrededor de las paradas representada por puntos o líneas.
- Opcionalmente, el nombre de la relación de salida (Luis Otavio Alvares, 2006).

La salida de las funciones *generaS* y *generaM* son respectivamente, una relación de paradas y una relación de movimientos. El parámetro *método* especifica cómo las paradas y los movimientos son calculados. Hasta el momento, dos métodos se han desarrollado: SMOT y CB-SMOT. SMOT es una técnica para encontrar paradas y movimiento, basado en la intersección. El principal inconveniente de esta hipótesis es que como el usuario es el que proporciona todas las paradas candidatas y la zona tope, los sitios más importantes que pueden llevar al descubrimiento de patrones interesantes se pueden perder sino son conocidos por el usuario. Se considera la intersección de una trayectoria con los tipos de rasgos especificados por el usuario para un período de tiempo mínimo (paradas candidatas).

En una trayectoria las paradas son los lugares importantes dónde el objeto se ha quedado por una cantidad mínima de tiempo. Siguiendo esto se genera un conjunto de datos (dataset), extraídos de los puntos de muestra de la trayectoria. Las paradas son interesantes localizaciones espaciales, también llamadas características espaciales.

Para encontrar diferentes paradas y movimientos se deben unir los puntos de muestra de la trayectoria con importantes características espaciales. Una técnica muy utilizada en este tipo de descubrimientos es la técnica SMOT.

Un movimiento se registra entre la parada anterior y la parada más reciente. La parada anterior puede ser nula debido a que la última parada pueda que sea la primera parada de la trayectoria (el inicio). Cuando un movimiento se inserta en el conjunto de los movimientos, sus características espacio-temporales también se agregan.

En esta técnica se explotan las funcionalidades de la BD espacial utilizando el exponente espacial disponible para comprobar si existe una parada de candidatos que se cruza con un vértice de la trayectoria. La descomposición de las trayectorias en paradas y movimientos proporciona acceso directo a



la información semántica de la trayectoria. Esta técnica fue propuesta basada en la idea de encontrar paradas en lugares candidatos.

Las paradas y los movimientos se calculan una sola vez, en una etapa previa del pre-procesamiento, y por lo tanto, el espacio de búsqueda espacial unido a la formulación de consultas se reducen al mínimo, en relación con el modelo de puntos de muestreo. La descomposición de las trayectorias en paradas y movimientos proporciona acceso directo a la información semántica de la trayectoria.

En palabras generales, la técnica verifica para cada punto de una trayectoria  $T$  si esta se corta con una parada candidata. En caso afirmativo, la técnica ve si la duración de la intersección es igual a un determinado umbral ya definido. En caso de que se cumpla la parada candidata se considera como una parada, y esta se registra. No se necesita almacenar los datos geográficos de la parada, ya que se almacena la instancia y el tipo de elemento espacial en la que se produce la parada.

La integración Apriori de las trayectorias que contienen información semántica geográfica que caracteriza a las partes más importantes de estas de acuerdo a la aplicación reduce de manera significativa la complejidad de las consultas a las BD y facilita el análisis de los datos de la trayectoria.

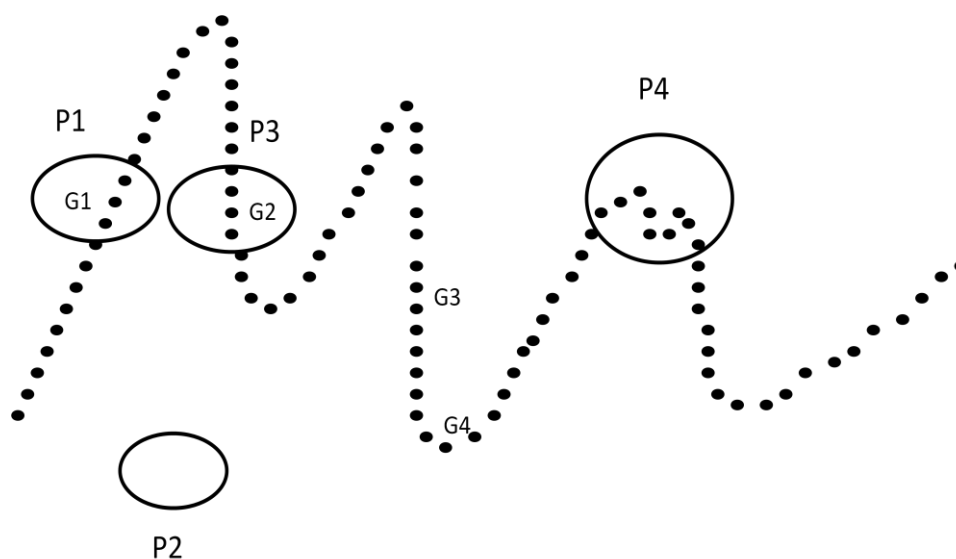
### **Técnica CB-SMOT**

La técnica CB-SMOT basada en la técnica SMOT identifica las paradas y los movimientos de una trayectoria. Este método trabaja con la velocidad para detectar de esta manera los lugares de interés en una trayectoria, por lo que se centra en encontrar regiones de baja velocidad. Esta técnica se apoya en otra llamada DBSCAN vista con anterioridad, al cual se le realizan algunas variaciones para su utilización en la implementación del CB-SMOT. La técnica DBSCAN debe contener en su concepto de vecindad los puntos sólo de la trayectoria que se está tratando, además debe considerar la distancia sobre la trayectoria y no la distancia directa entre dos puntos.

En la primera etapa las partes más lentas de una trayectoria son las posibles paradas, estas son identificadas mediante la variación de la técnica DBSCAN que considera la línea unidimensional (trayectorias) y la velocidad. En el segundo paso, la técnica de identificación es donde estas paradas

potenciales que se ven en la primera etapa se encuentran, teniendo en cuenta la geografía detrás de las trayectorias.

CB-SMOT toma de las paradas candidatas entradas por el usuario, cada parada potencial y las pruebas tanto de intersección como la duración mínima. Si las paradas potenciales no se cruzan con ninguna parada candidata aún estas pueden ser consideradas lugares interesantes. Para proporcionarle esta información al usuario la técnica etiqueta estos lugares como paradas desconocidas.



**Figura 1:** Ejemplo de trayectoria con paradas desconocidas

En la Figura 1 se muestran cuatro paradas y 4 grupos (G1, G2, G3 y G4). Las paradas candidatas identificadas por el usuario son: P1, P2, P3 y P4. El grupo G1 se cruza con la parada candidata P1 por lo que la primera parada de la trayectoria es P1, al igual que el grupo G2 que se cruza con la parada P3 por lo que esta constituye la segunda parada de la trayectoria. Los grupos G3 y G4 no se cruzan con ninguna parada candidata por lo que las otras paradas se consideran paradas desconocidas.

Cada parada que sea considerada desconocida va a recibir un identificador, y si alguna de estas se cruzan recibirán el mismo. Un ejemplo de esto puede verse evidenciado en la Figura 2. Existe una trayectoria T1 la cual posee una parada candidata P1 y un grupo G1, este grupo no se corta con la parada

candidata. De igual manera se presenta una trayectoria T2 que posee una parada candidata P2 y un grupo G2 que no se corta con esta parada candidata. Estas son consideradas como paradas desconocidas. Como los grupos G1 y G2 se cortan estas dos paradas recibirán el mismo identificador.

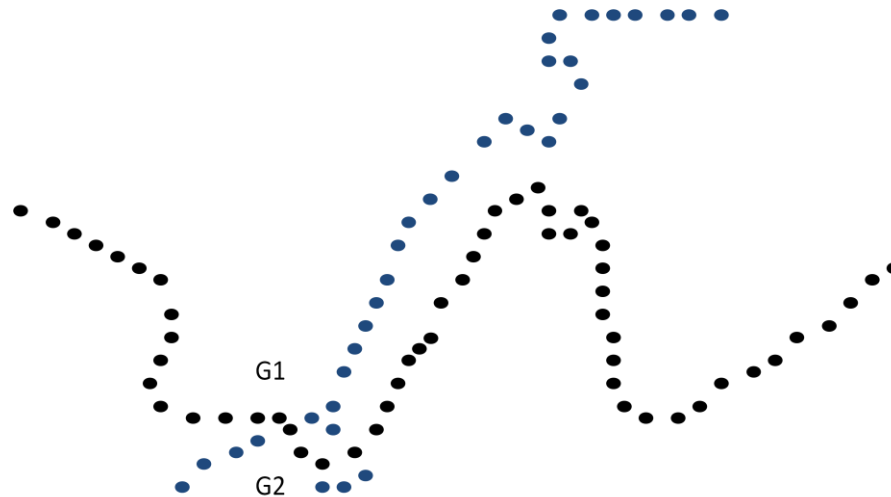


Figura 2: Dos trayectorias con paradas desconocidas

Para el método de CB-SMOT el usuario proporciona los mismos parámetros que para el método SMOT, pero además de estos proporcionan un valor extra que es necesario para encontrar grupos en las trayectorias individuales, el parámetro *Eps*. El parámetro *Eps* indica la distancia absoluta utilizada para calcular el entorno de un punto. Sin embargo, le es difícil al usuario especificar un buen valor para este parámetro debido a que no conoce bien las características de cada trayectoria y el valor que se brinda es absoluto. Para esto se elabora una alternativa para el usuario con el fin de ajustar este parámetro (Andrey Tietbohl, 2008).

Este método es apropiado para analizar casos donde la velocidad sea importante, por esto su implementación se basa en este elemento. Gracias al apoyo que posee en la técnica DBSCAN, su funcionamiento es apropiado para trabajar con situaciones de análisis de tráfico.

### **2.4 Aplicaciones de las técnicas para el control de flotas**

Las técnicas planteadas y descritas anteriormente se encuentran capacitadas para desarrollar análisis sobre las trayectorias de los objetos en movimiento. Brindan la posibilidad de conocer información sobre entidades móviles que a través de simples técnicas estadísticas no puede ser extraída de la BD.

La utilización de estas técnicas en el desarrollo de las aplicaciones basadas en el control de flotas contribuirá a reducir trabajo y mano de obra innecesaria debido a que brindan posibilidades como: la predicción de movimientos futuros, el seguimiento y control de entidades móviles, rápida manipulación de los datos e información extraída de una trayectoria, entre otras ventajas que constituyen en gran paso de avance en el perfeccionamiento de los SIG.

### **2.5 Conclusiones Parciales**

En el capítulo se han caracterizado técnicas de MD espacio-temporal con aplicaciones para el control de flotas brindando base y argumentos para la confirmación de que el empleo de las mismas en los SIG constituiría un gran avance. Las técnicas descritas son capaces de detectar características importantes en las trayectorias por lo que brindan útiles resultados extraídos a partir del análisis realizado sobre ellas. Se brindaron características importantes de las regiones de interés que constituyen una parte esencial en el proceso de MD espacio-temporal facilitando la comprensión de su utilidad en el desarrollo de algunas técnicas. La realización del capítulo permite poseer elementos para lograr comprender profundamente el funcionamiento de cada técnica además de, aportar importantes detalles para realizar la comparación.

### CAPÍTULO 3: COMPARACIÓN Y SELECCIÓN DE LAS TÉCNICAS

#### 3.1 Introducción

En el presente capítulo se estarán tratando aspectos relacionados con el funcionamiento de las técnicas abordadas en el capítulo dos epígrafe 2.2. Se describen resultados de experimentos realizados a dos técnicas, obteniendo de esta forma argumentos que complementan los resultados de la selección realizada. En este capítulo se evidencia además, características propias del funcionamiento de cada técnica y resultados obtenidos por experimentos realizados por diversos autores, logrado así obtener recursos importantes que serán utilizados en la comparación. Se definen criterios con el objetivo de lograr una selección de técnicas idóneas a incorporar en los SIG para el control de flotas.

#### 3.2 Análisis del funcionamiento de las técnicas.

Las diferentes técnicas tratadas presentan diversas características que permiten la extracción de conocimiento a partir de datos espacio-temporales almacenados en bases de datos. Estas técnicas son implementadas de acuerdo a los requisitos que se necesitan cubrir y a las metas que deben cumplir. Son probadas de disímiles formas con el objetivo de valorar los distintos resultados y observar así su comportamiento ante diferentes entradas de datos. La presencia de dificultades en algunas puede constituir desventajas en cuanto a parámetros de evaluación. A continuación se describirá el funcionamiento de las técnicas de MD espacio-temporal y se hará alusión a experimentos desarrollados por diferentes autores.

#### Técnica Periódica

La técnica periódica está compuesta por dos partes explicadas en el epígrafe 2.3.1, se muestran a continuación funcionalidades realizadas en ellas:

- En la primera se definen por primera vez todos los puntos de referencia y para cada punto de referencia se detectan los períodos.

- En la segunda se examinan los puntos de referencia con el período correspondiente para realizar la extracción de los comportamientos periódicos los cuales pueden ser definidos como la repetición de actividades en lugares determinados en intervalos de tiempo regulares.

El período de tiempo siempre va a ser más fácil de detectar si se consideran los datos del estudio realizado al lugar o a la región de interés. La transformación de Fourier no permite detectar con claridad los períodos correctos porque este método es sensible al ruido espacial, por lo que se hace necesario analizar las acciones realizadas periódicamente por el objeto en movimiento. Una vez detectados los comportamientos periódicos, estos deben ser agrupados y analizados para obtener el número óptimo. La segunda parte de la técnica periódica es un método jerárquico que se basa en determinar la cantidad de comportamientos periódicos.

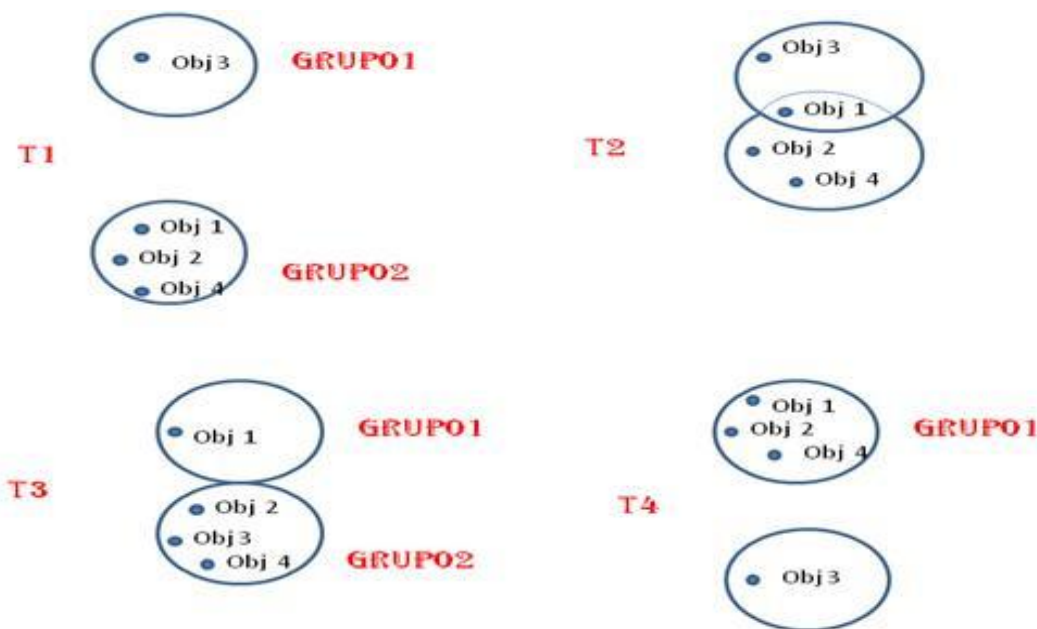
Los experimentos realizados a la técnica en (Zhenhui Li, 2007) se observaron los resultados que se obtuvieron de las diferentes pruebas realizadas, se muestra la eficiencia de la técnica en cuanto a la capacidad de filtrar las redundancias entre los segmentos que son generados por comportamientos periódicos, logrando de esta forma obtener los verdaderos comportamientos.

### **Técnica ObjectGrowth**

Esta técnica está diseñada para la eficiente extracción de enjambres cerrados (se explica en el epígrafe 2.3.1). Utiliza dos reglas de poda para acelerar el proceso de búsqueda: la primera llamada Apriori poda para detener la estructura cuando se presenta un recorrido que no puede satisfacer los instantes de tiempo y la segunda es la llamada con versiones anteriores de poda. La última regla se utiliza para aplicar la propiedad de cierre donde se comprueba si otro súper-conjunto posee una mayor cantidad de objetos y la misma o menos cantidad de instantes de tiempo que el conjunto que se está trabajando en ese momento, por lo que pasaría a ser un enjambre cerrado. El paso de comprobación de cierre se realiza en el proceso de búsqueda. Se debe tener en cuenta que para que un conjunto sea un enjambre cerrado debe cumplir con las condiciones de los parámetros *Min<sub>o</sub>* (cantidad mínima de objetos dentro del conjunto) y *Min<sub>t</sub>* (cantidad mínima de instantes de tiempo).

ObjectGrowth aplica la técnica DFS para realizar la búsqueda de los enjambres cerrados. Inicialmente la técnica comienza a evaluar para el primer objeto que se encuentra en el espacio de búsqueda. Realiza

todas las combinaciones y refleja los instantes de tiempo en los que se encuentra cada una de estas hasta encontrar el enjambre cerrado. Para esto se van comprobando con las dos reglas de poda en cada iteración realizada. Este mismo procedimiento se repite para cada uno de los restantes objetos del espacio de búsqueda. (Figura 3 y Figura 4).



**Figura 3:** Agrupaciones de objetos en diferentes instantes de tiempo

Según (Song Li, 2009) en experimentos realizados a la técnica ObjectGrowth demuestra una mejor eficiencia que las técnicas anteriores. Esto se debe a que la técnica se centra en la búsqueda de enjambres cerrados mientras que los anteriores extraían todos los enjambres presentes en el espacio de búsqueda. Además también tiene grandes ventajas en cuanto al tiempo de ejecución; las versiones anteriores trabajaban la segunda regla de poda de manera más estricta (Zhenhui Li, 2009) y no lo realizaban en el proceso de búsqueda sino en una etapa de pre-procesamiento, esto trae como consecuencia que su tiempo de ejecución sea mucho más elevado que el de ObjectGrowth.

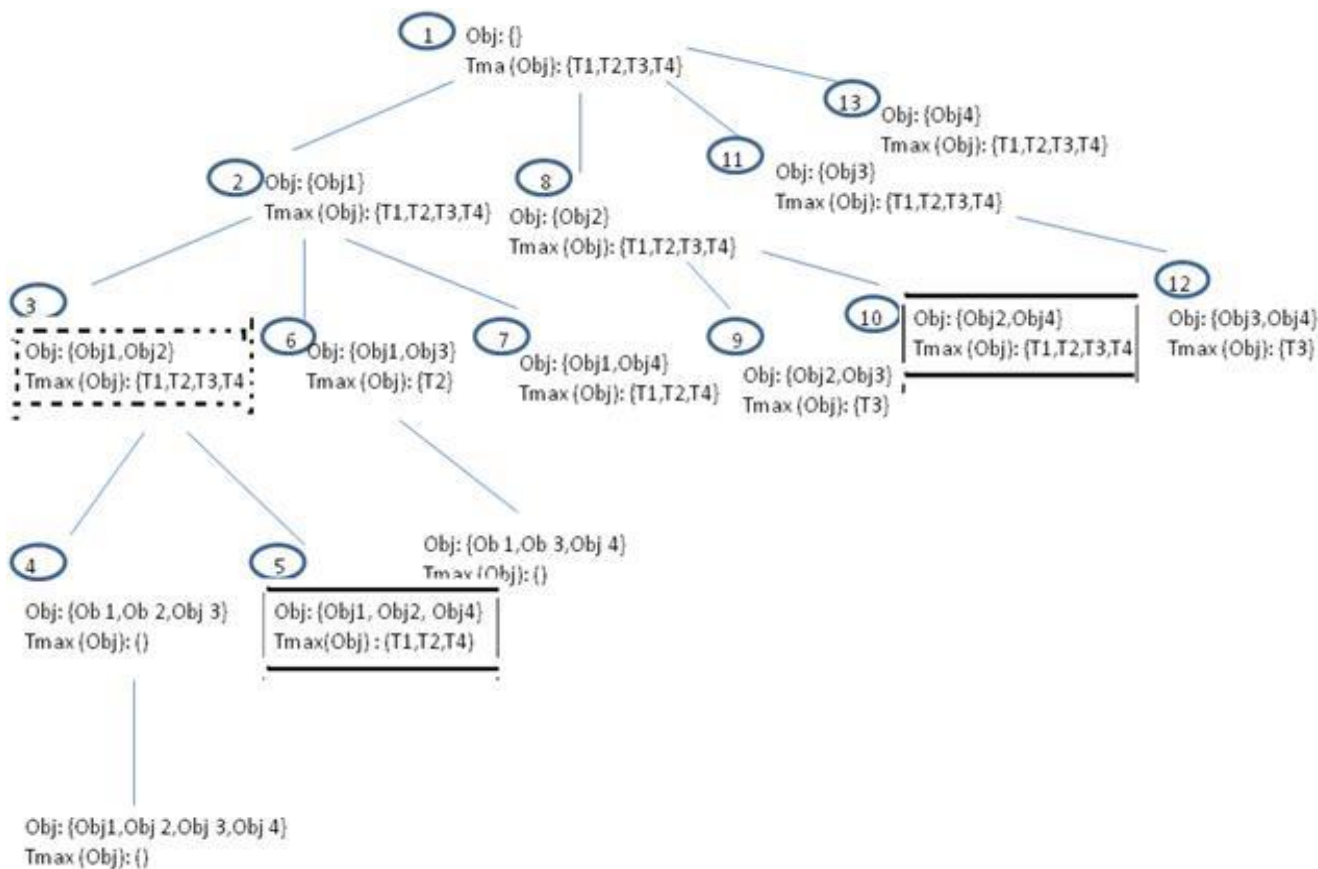


Figura 4: Técnica ObjectGrowth

### Técnica DBSCAN

En la actualidad existen diversos métodos de agrupamiento que realizan y aplican diversos parámetros para lograr efectividad en su realización. La técnica DBSCAN tiene como objetivo la identificación de clases, es decir, la agrupación de los objetos de una BD en sub-clases significativas. Se deben tener en cuenta para esta técnica una serie de requisitos como: poseer conocimientos del tema para determinar los parámetros de entrada porque los valores apropiados a menudo no se conocen de antemano cuando se trata de grandes BD.

Los métodos de agrupamientos conocidos no combinan esta serie de requisitos. Para esto se propone la técnica DBSCAN. La idea esencial de esta técnica es que cada punto de un grupo de la zona de un radio



determinado tiene que contener al menos un número mínimo de puntos (*MinPts*), es decir, la densidad en la zona tiene que superar cierto umbral. El vecindario está determinado por la elección de una función de distancia de dos puntos  $p$  y  $q$ , denotado por  $dist(p, q)$  (Martin Ester, 1996).

La densidad accesible es simétrica para pares de puntos básicos, sin embargo, no es simétrica cuando se encuentran involucrados un punto de la frontera con un punto central. Esto se debe a que desde un punto de la frontera no se puede acceder a un punto central, y desde un punto central se puede acceder a un punto de la frontera o a un punto central, dos puntos fronterizos no pueden accederse por lo que se crea una conexión con otro punto central (Figura 5).

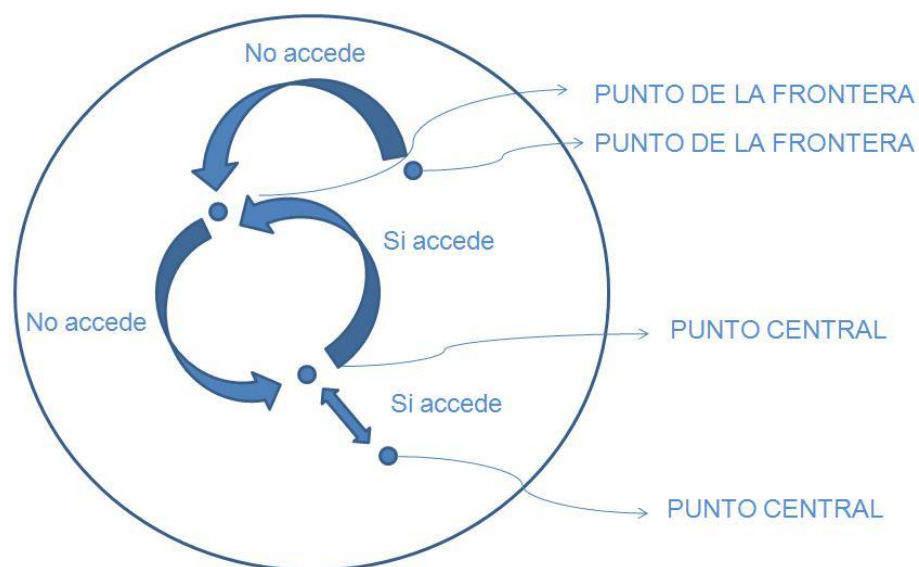


Figura 5: Densidad accesible

DBSCAN está diseñado para descubrir los grupos y el ruido en una BD espacial (Figura 6). Para esto habría que conocer los parámetros *Eps* y *MinPts* de cada grupo y por lo menos un punto de la agrupación respectiva. Para el cálculo de estos parámetros se desarrolla una eficiente heurística (Martin Ester, 1996) de los agrupamientos menos poblados. Utilizando los parámetros correctos se pueden obtener todos los puntos de densidad accesible desde un punto determinado.

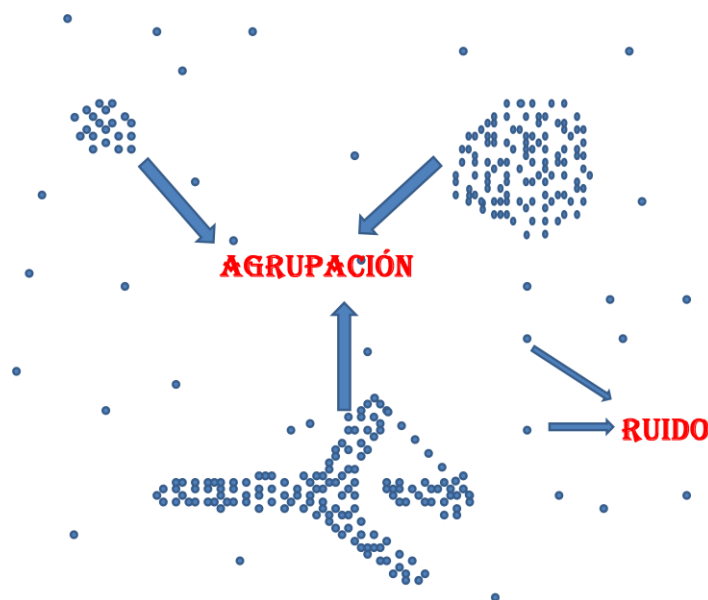


Figura 6: Agrupaciones y ruidos (DBSCAN)

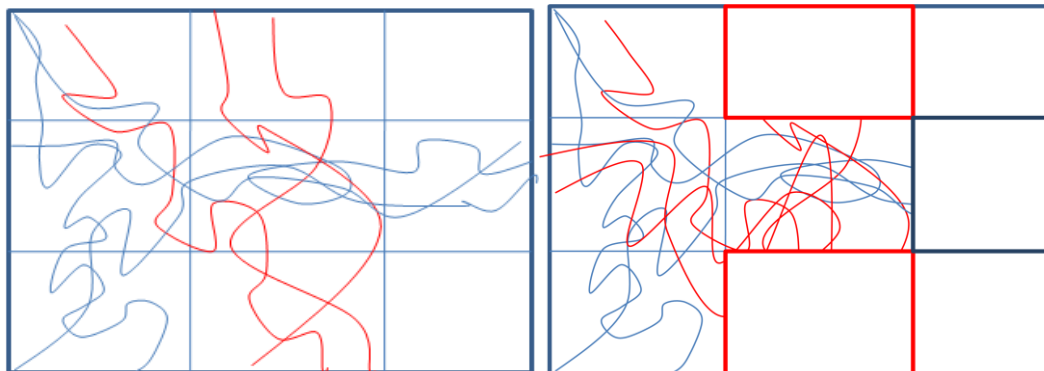
En experimentos realizados a esta técnica según (Martin Ester, 1996) se trabajó con datos reales referentes a polígonos, tramos y puntos. A la técnica se le realizaron comparaciones con respecto a otras de igual magnitud y con objetivos similares. Luego de analizar los resultados arrojados por todos los algoritmos probados se llegó a la conclusión de que la técnica DBSCAN contiene importantes ventajas para su aplicación en el control de flotas. Esta técnica presenta facilidades en su implementación que le permiten detectar e identificar el ruido o perturbaciones presentes en los datos de las distintas agrupaciones que va conformando, de ahí la capacidad de poder distinguir el trabajo con datos de relevante importancia y con otros que no son de gran valor para las aplicaciones.

## Técnica Traclclass

La técnica Traclclass realiza agrupaciones basadas en trayectorias y en regiones. (Ver Figura 7 y Figura 8). A continuación se enumeran un conjunto de funcionalidades realizadas por Traclclass:

1. El agrupamiento basado en regiones descubre las regiones que tienen trayectorias en su mayoría de una clase, independientemente de los patrones de movimiento.
2. El agrupamiento basada en trayectorias descubre sub-trayectorias que indican los patrones

comunes de movimiento de cada clase. La colaboración entre los dos tipos de agrupamiento conduce al descubrimiento de dos tipos de características discriminatorias, aumentando la precisión de clasificación de manera significativa.

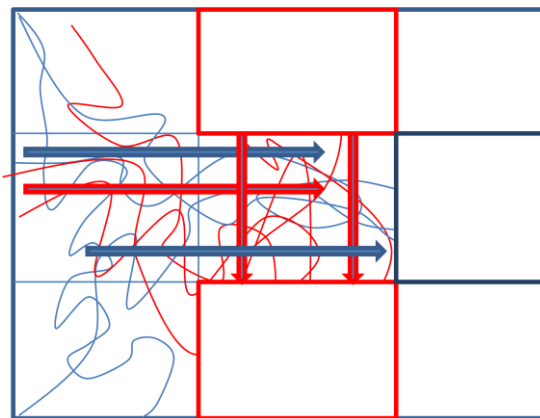


**Figura 7:** Agrupamiento basado en regiones

Una tarea esencial para la clasificación efectiva es la generación de características discriminatorias (características no comunes en todas las trayectorias). Las características discriminatorias pueden aparecer en partes de las trayectorias, pues no sólo aparecen como patrones de movimiento común sino también en el agrupamiento basado en regiones. Las trayectorias inicialmente deben ser particionadas, formalizándose así el problema de la agrupación basada en regiones usando el principio de descripción de longitud mínima (MDL) y un eficiente algoritmo de aproximación de complejidad  $O(n)$  para encontrar las particiones más óptimas posible (Jae-Gil Lee, 2008).

El procedimiento de agrupamiento basado en trayectoria se centra en el sistema de particionado y agrupamiento (Jae-Gil Lee, 2008). El sistema de particionado y agrupamiento se amplía con el propósito de incorporar las etiquetas de las clases a los agrupamientos. Para que la técnica pueda comenzar a desarrollarse, las trayectorias deben estar clasificadas, es decir, pertenecer a un grupo de clases. Luego se procede a realizar el agrupamiento basado en regiones. Las regiones que presentan una clase son mayormente homogéneas, por tanto las que contengan la mayoría de las trayectorias de una misma clase se dice que son regiones homogéneas. Las regiones no homogéneas, es decir la que tienen trayectorias

de varias clases, son cuantificadas recursivamente por la técnica para encontrar más regiones homogéneas. Estas regiones homogéneas que van siendo encontradas se utilizan de agrupamientos basados en regiones.



**Figura 8:** Agrupamiento basado en trayectorias

Cuando no se encuentren más regiones homogéneas en las no homogéneas, se pasa a aplicar la otra parte de la técnica para encontrar agrupamientos basados en trayectoria. Esta parte de la técnica se basa en descubrir patrones de movimiento comunes de cada clase en las regiones no homogéneas. Los patrones que se encuentren en la trayectoria de un tipo de clase y no en la trayectoria del otros tipos de clases se dice que son patrones discriminatorios, es decir, estos son encontrados solamente en la trayectoria de un tipo de clase, no son comunes para las todas clases. Luego de haber encontrado estos, se vuelven a analizar los patrones no discriminatorios para encontrar más patrones discriminatorios. Los patrones discriminatorios encontrados son tomados de agrupamientos basados en trayectoria. El uso de estos dos agrupamientos permite detectar más características de alta calidad.

El agrupamiento basado en regiones descubre características de mayor nivel que el agrupamiento basado en trayectorias debido a que no utiliza patrones de movimiento. Además en cada agrupación va a existir una jerarquía desde los grupos más grandes a los grupos más pequeños, por lo que las características van siendo encontradas de arriba hacia abajo. Las características de los niveles más altos son más importantes que las características de los niveles más bajos. Después de haber realizado el agrupamiento basado en trayectorias, Traclass selecciona los que tengan un alto poder discriminatorio para su uso en la

clasificación. Después de realizar todos los agrupamientos se utilizan métodos que refinan y hacen más efectiva la clasificación de las trayectorias. (Figura 9)

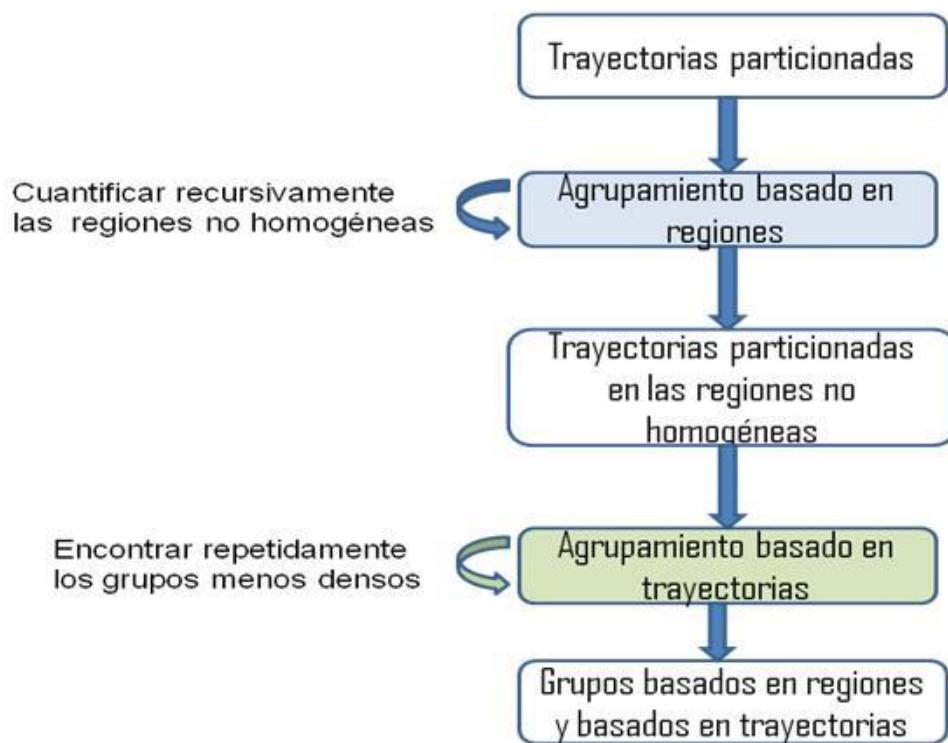


Figura 9: Técnica Traclass

Según (Jae-Gil Lee, 2008) en la realización de experimentos a la técnica Traclass se tiene en cuenta la aplicación de los agrupamientos en su conjunto (versión completa) y del agrupamiento basado en regiones individualmente (versión simple). En estos se evidencian tres parámetros fundamentales: la exactitud de la clasificación, el tiempo de predicción y el tiempo de entrenamiento, el cual no es tomado en cuenta debido a que para el control de flotas no representa un requisito primordial. La efectividad de la clasificación es medida de acuerdo al total de trayectorias clasificadas correctamente del total de trayectorias de prueba que se tienen.

Se introduce un juego de datos (Jae-Gil Lee, 2008) que cuenta con información acerca del movimiento animal con tres tipos de clases, la trayectoria de los huracanes con cinco tipos de clases y un buque de

navegación con dos tipos de clases. Se demostró que la versión simple es inferior totalmente en cuanto a efectividad que la versión completa (Figura 10), ya que esta omite características importantes para la clasificación de las trayectorias. La alta eficiencia de la versión completa es lograda debido a la realización de los dos agrupamientos, esto a pesar de que el segundo sea más costoso por el trabajo que realiza con los patrones de movimiento.

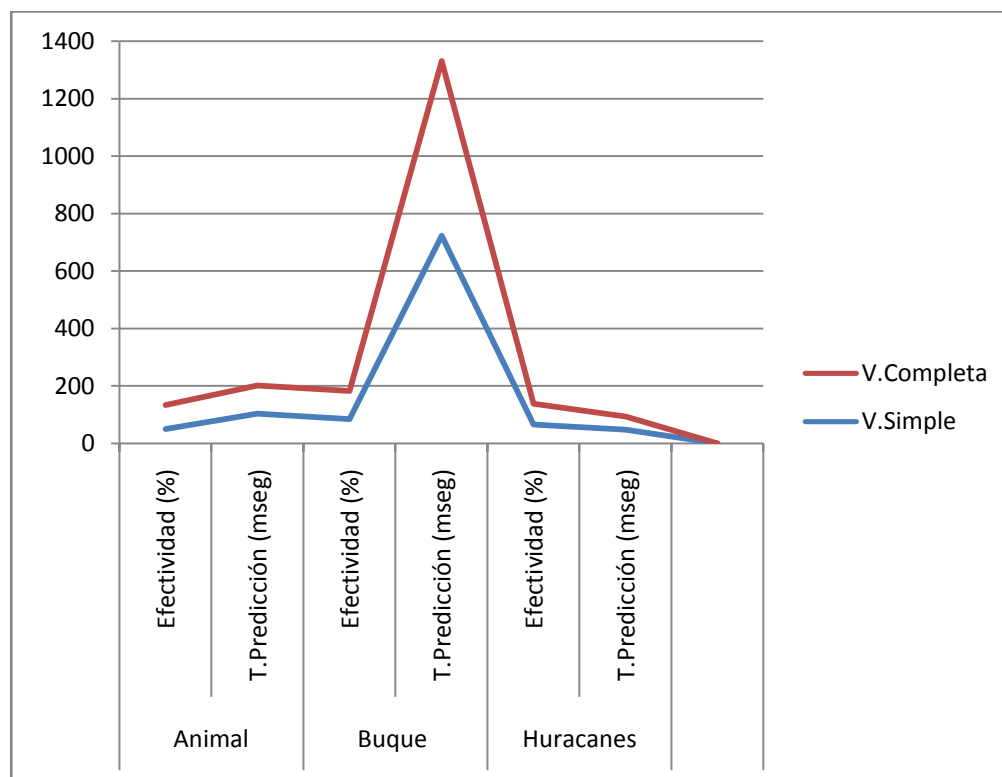


Figura 10: Comparación de versiones

## Técnica Traclus

Traclus realiza el agrupamiento de un conjunto de objetos físicos o abstractos en clases de objetos similares, además del análisis de datos típicos encontrando así los objetos que se han movido de una manera similar. En ocasiones, cuando se agrupan las trayectorias en su conjunto, no se puede evidenciar comportamientos comunes en ellas, ya que cada una toma direcciones diferentes; sin embargo, separándolas en tramos, se puede observar que en algunas partes presentan comportamientos comunes.

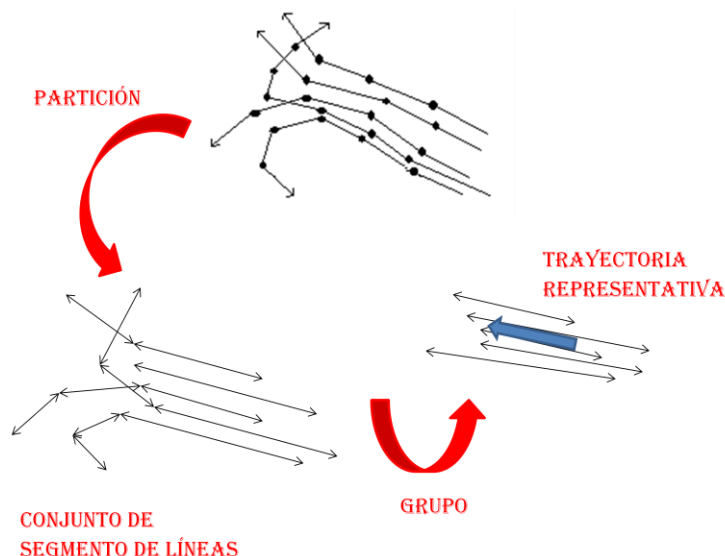
De esta forma se puede perder, en ocasiones, información valiosa si se tratan las trayectorias como un todo. Una solución para esto es dividir la trayectoria en un conjunto de segmentos de línea y luego agrupar los segmentos de líneas similares. Este sistema es llamado sistema de partición y grupo. La principal ventaja de este sistema es el descubrimiento de sub-trayectorias comunes a partir de una BD de la trayectoria.

El sistema de particionado y agrupamiento está compuesto por las siguientes fases:

1. Fase de particionado: Cada trayectoria es óptimamente dividida en segmentos de líneas, y estos segmentos de líneas pasan a la otra fase.
2. Fase de agrupamiento: Los segmentos de líneas similares se agrupan. Se utilizan métodos basados sobre densidad específicos ya que pueden descubrir los grupos de forma arbitraria y pueden filtrar los ruidos.

La técnica Traclus (Figura 11) es desarrollada de acuerdo al sistema de particionado y agrupamiento. En esta técnica una trayectoria puede pertenecer a varios grupos, ya que primero se realiza el particionado y agrupamiento de la trayectoria en segmentos de línea y luego se hacen los grupos, de estos segmentos de línea. Luego realiza una trayectoria representativa, la cual es una secuencia de puntos al igual que en una trayectoria común. Es una trayectoria imaginaria que indica el comportamiento de los principales segmentos de línea dentro del grupo.

El procedimiento general de esta técnica es primeramente particionar las trayectorias en sub-trayectorias (segmentos de línea), luego las sub-trayectorias que se encuentran cerca una de la otra de acuerdo a la medida de distancia planificada se agrupan y por último se genera una trayectoria representativa por cada grupo. Para particionar la trayectoria se propone un algoritmo de particionado de la trayectoria el cual tiene presente dos propiedades importantes: precisión y concisión (Jae-Gil Lee, 2007). Esta técnica parte del supuesto de que todos los segmentos de línea están clasificados. A medida que la técnica va avanzando estos se van clasificando como un grupo o como ruido.



**Figura 11:** Técnica Traclus

Para esta técnica se utilizan dos parámetros de entrada (Jae-Gil Lee, 2007) los cuales pueden ser insertados por el usuario o calculados. Con el objetivo de utilizar esta técnica para el control de flotas la misma debe ser aplicada con los valores óptimos de estos parámetros, para de esta forma asegurar un mejor manejo de las trayectorias con las que se estén trabajando. Basándose en lo expuesto por Lee, 2007, para realizar un conjunto de experimentos se tomaron datos sobre el movimiento animal y el seguimiento de huracanes. De acuerdo a estos experimentos realizados a la técnica queda demostrado que la misma realiza un mejor particionado, agrupamiento y representación de las trayectorias representativas cuando se utilizan los valores óptimos de los parámetros.

Esta técnica posee una gran ventaja para el control de flotas ya que trabaja con conjuntos de sub-trayectorias similares, por lo que las características observadas en las mismas pueden ser de gran ayuda en diversas esferas que tributen a aplicaciones destinadas a este propósito.

## **Técnica DFS\_MINE**

La técnica DFS\_MINE presenta como objetivo encontrar secuencias de eventos frecuentes en datos espacio-temporales, además de realizar un rápido descubrimiento de patrones secuenciales en la BD. La



técnica busca en la BD los elementos frecuentes de acuerdo al parámetro `min_support` entrado por el usuario, el cual se utiliza para definir la cantidad de veces que debe estar el elemento en la BD y se almacenan en la lista `MaxFreqList` (la lista de un máximo de secuencias frecuentes). Luego con los elementos de esta lista se forman secuencias de longitud dos realizando todas las combinaciones posibles de los elementos. Una vez que son formadas las secuencias se compara nuevamente con la BD para reafirmar la frecuencia de las mismas y las que sean no frecuentes pasan a la lista `MinFreqList` (la lista de un mínimo de secuencias no frecuentes), la que se utilizará para desechar luego combinaciones no frecuentes. De esta forma también se determina el conjunto `Useless` que representará las frecuencias de longitud  $n$  que no se interceptarán con las demás. El proceso se repetirá hasta conformar las secuencias frecuentes más largas.

Antes de insertar una secuencia frecuente en la lista `MaxFreqList` se debe tener en cuenta que la misma no puede tener una súper-secuencia, esto se debe a que la técnica cuando descubre una secuencia  $k$  frecuente no agota recursos en la búsqueda de las secuencias  $k-1$ . Todas las sub-secuencias de la secuencia entrada deben ser frecuentes. Una vez insertada la secuencia a la lista serán eliminadas todas sus sub-secuencias de la lista. El proceso llevado en la lista `MinFreqList` es similar lo que con las sub-secuencias no frecuentes.

En algunos experimentos realizados a la técnica en (Ilias Tsoukatos, 2001) se pudo evidenciar que esta presenta gran ventaja en cuanto a espacio en memoria, ya que no requiere de una gran capacidad para guardar información referente a las secuencias. Esta técnica en comparación con otras es un poco más lenta aunque a medida que aumenta la cantidad de elementos el tiempo sólo aumenta en un porcentaje muy bajo, al contrario de las otras técnica que en ocasiones se quedan sin memoria a utilizar debido a la gran cantidad de información que almacenan de las frecuencias, y esta aumenta considerablemente con el aumento de los elementos.

### **Técnicas SMOT y CB-SMOT**

Las técnicas SMOT y CB-SMOT son similares en cuanto a su funcionamiento, la diferencia es que CB-SMOT se basa en la velocidad, por eso es aconsejable utilizarla en aplicaciones donde la velocidad sea

un parámetro de gran importancia. CB-SMOT posee, a diferencia de SMOT, la entrada del parámetro *Eps*, el cual es utilizado en la variante de la técnica DBSCAN que es aplicada en CB-SMOT.

La idea esencial de estas técnicas es encontrar las paradas y los movimientos de una trayectoria, es decir, tratarla semánticamente, logrando así mantener un estricto control de la entidad móvil. Por esto son de gran utilidad ambas en las aplicaciones para el control de flotas.

### 3.3 Experimentos realizados

Se realizan experimentos con el objetivo de poseer mayores argumentos que puedan resultar de gran valor en la comparación realizada a las técnicas estudiadas; estos brindarán evidencias que se pueden utilizar para la toma de decisiones en la selección.

Las técnicas ObjectGrowth y Traclus descritas en el epígrafe 2.2 se encuentran implementadas en la página Movemine, visualizando la trayectoria recorrida por diferentes especies de animales. En un experimento realizado a 10 individuos de águilas calvas en movemine la técnica Traclus representó en un tiempo de 78 milisegundos lo mostrado en Figura 12. Traclus encontró 16 agrupaciones para un total de 7334 segmentos de línea. La distancia mínima especificada fue de 0.8 y la cantidad mínima de puntos de cuatro.

Con la técnica ObjectGrowth también se realizaron experimentos en la página movemine. Para esta técnica la selección contaba con 43 individuos de la especie de Swainsoni (especie que pertenece al halcón) escogida en un intervalo de tiempo del 29 de julio de 1995 al 24 de junio de 1998 mostrado en la Figura 13. En un tiempo de 156 milisegundos la técnica encontró 4 enjambres, demorando sólo 62 milisegundos en encontrar todos los enjambres cerrados.



Figura 12: Visualización de la técnica Tracrus



Figura 13: Visualización de la especie Swainsoni

### 3.4 Comparación y selección de las técnicas.

Con la aplicación de criterios definidos de acuerdo al funcionamiento de las técnicas para realizar la comparación entre las mismas, se logrará efectuar una óptima selección de las más idóneas para el control de flotas. Para esto primeramente se agruparán las técnicas de acuerdo a la función que realizan y luego estas serán comparadas utilizando criterios comunes y de acuerdo a sus funciones. A continuación se evidencian los criterios a utilizar:

- **Eficiencia:** Las técnicas a utilizar deben realizar un profundo análisis de los datos espacio-temporales logrando así convertirlos en información que tribute al conocimiento utilizado en la toma de decisiones.
- **Velocidad:** La velocidad de ejecución de cada técnica puede presentarse como una ventaja. Esta podría posibilitar un mayor nivel de aceptación debido que la rapidez es un parámetro siempre a tener en cuenta en aplicaciones dedicadas al control de una entidad móvil.
- **Nivel de precisión:** La precisión de los valores arrojados es muy importante en aplicaciones para el control de flotas. Es necesario trabajar con resultados específicos, lógicos y esperados, logrando así que la veracidad y eficiencia del resultado final posea un nivel elevado.

Los criterios planteados están estrechamente relacionados ya que algunos son la base para poder cumplir otros. Un ejemplo de esto se evidencia en la efectividad, siendo esta necesaria en la técnica para lograr obtener resultados con un mayor nivel de precisión. La selección de técnicas que cumplan con la mayoría de estos criterios será un gran paso de avance ya que posibilitará que en los SIG para el control de flotas sean incorporadas las más efectivas, precisas y rápidas.

#### 3.4.1 Agrupamiento de las técnicas de acuerdo a su funcionamiento

Las técnicas vistas en la investigación están implementadas para diferentes propósitos; esto provoca que el funcionamiento de las mismas sea de acuerdo al objetivo para el que están destinadas. De ahí la necesidad de agrupar cada una en distintas funcionalidades.

**Tabla 1:** Agrupamiento de las técnicas basado en sus funcionalidades.

<b>Análisis de Trayectorias</b>	<b>Técnicas</b>
<b>Trabajo con sub-trayectorias comunes</b>	Traclass, Traclus
<b>Análisis semántico</b>	SMOT, CB-SMOT
<b>Detección de comportamientos grupales</b>	ObjectGrowth
<b>Agrupamiento</b>	DBSCAN
<b>Predicción de movimientos futuros</b>	<b>Técnicas</b>
<b>Detección de patrones frecuentes</b>	DFS_Mine, Periódica

### 3.4.2 Comparación y solución

Luego de haberse realizado un minucioso y detallado trabajo con las técnicas, éstas deben ser comparadas para lograr que la selección cumpla con las expectativas propuestas. Se debe alcanzar un gran avance con la incorporación de las técnicas seleccionadas en los SIG, debido a que las mismas brindarán la posibilidad de que contengan razonamiento, siendo capaces así de realizar diversas tareas que anteriormente no ejecutaban.

#### Técnicas SMOT y CB-SMOT

En experimentos realizados en (Andrey Tietbohl, 2008) se brindan comparaciones entre las técnicas SMOT y CB-SMOT, los cuales son de gran importancia para el control del tráfico. La técnica CB-SMOT, para una entrada de datos reales donde las paradas candidatas no cubren la región completa, encuentra una gran cantidad de paradas desconocidas, mientras que la técnica SMOT encuentra pocas paradas conocidas, siendo aplicadas ambas a la misma área. Esto se pone de manifiesto cuando las paradas candidatas entradas por el usuario no cubren toda el área de trabajo.

En caso de que las paradas candidatas cubran toda la región, la técnica SMOT encuentra disímiles paradas conocidas, mientras que la técnica CB-SMOT encuentra menos paradas conocidas. La

explicación está basada en la diferencia en cuanto al desarrollo de ambas técnicas; lo que resulta una parada para la técnica SMOT puede ser que no sea así para CB-SMOT ya que esta se basa en la velocidad, de ahí la importancia de su utilización en aplicaciones donde se considera la velocidad como parámetro de análisis.

En muchas aplicaciones como, por ejemplo, la de gestión del tráfico y el turismo, la técnica CB-SMOT puede encontrar varios lugares de interés que son desconocidos por el usuario. En una aplicación de gestión del tráfico CB-SMOT encontraría rotondas, semáforos y controladores de velocidad, incluso si no se dan como paradas candidatas por el usuario. En una aplicación de turismo, CB-SMOT puede encontrar varios lugares inesperados y aunque esta aplicación no tenga nada que ver con el turismo puede que estos lugares atraigan a los turistas. Aunque SMOT sólo puede descubrir paradas relacionados con los lugares interesantes definidos concretamente por el usuario, puede brindar mejores resultados que el CB-SMOT en aplicaciones donde la velocidad no es relevante.

Una observación importante sobre el CB-SMOT es la capacidad que posee para generar grupos no existentes en algunas partes de la trayectoria. Esto es muy común dentro de los edificios, donde puede ser que la señal de los GPS insertados en los celulares no llegue. Por lo general, en este caso, los puntos de muestreo de la trayectoria se interrumpen en el momento en que el objeto en movimiento entra en un edificio y se reiniciará cuando el objeto sale del edificio. Las técnicas de agrupaciones basadas en densidad no son capaces de encontrar directamente grupos en esas regiones; pero esta técnica va a descubrir lugares, porque la velocidad media entre los dos puntos (entrada y salida de un edificio) es muy baja.

Teniendo en cuenta todo lo visto en cuanto al funcionamiento de las técnicas SMOT y CB-SMOT, se considera según los criterios definidos con anterioridad en el epígrafe 3.2.1 que CB-SMOT posee una mayor eficiencia que la técnica SMOT. CB-SMOT facilita disímiles ventajas al ser utilizada ya que brinda la posibilidad de obtener lugares de referencia que no hayan sido especificados por el usuario e identificar paradas en las zonas donde la señal de los GPS no llegan; esto permite que los resultados del análisis semántico de la trayectoria realizado sean lógicos y esperados, de ahí que a su vez el nivel de precisión de CB-SMOT sea más elevado. Otra de las características que posee CB-SMOT es que su trabajo en lugares donde el parámetro *velocidad* sea importante es más eficaz que el realizado por la técnica SMOT,

lo que resulta de gran ayuda en aplicaciones para el control de flotas. La técnica CB-SMOT actúa de manera más rápida que SMOT debido a que el análisis sobre las trayectorias que realiza es menos riguroso porque no se basa en la velocidad. La selección propuesta para lograr una mayor efectividad, rapidez y precisión es la técnica CB-SMOT.

### **Técnicas Traclass y Traclus**

Las técnicas Traclass y Traclus poseen ventajas en cuanto a la clasificación de las trayectorias trabajando con sus particiones. Traclass a medida que trabaja con las trayectorias encuentra más características que Traclus debido a que utiliza dos tipos de agrupamiento vistos en el capítulo 3 epígrafe 3.2. Se basa además en el sistema de particionado y agrupamiento, funcionamiento que también lleva a cabo Traclus. Las características aportadas por esta técnica pueden brindar una mayor cantidad de datos a la clasificación de las trayectorias.

Basándose en el criterio de eficiencia se puede considerar que la técnica Traclass posee un nivel más elevado que Traclus. Traclass demora mucho más tiempo ejecutándose debido al número de algoritmos que en ella se desarrollan, pero la efectividad es mucho más notable al aportar la mayor cantidad de características esenciales y que identifican a cada una de las trayectorias para su clasificación. Traclus más bien se centra en el particionado y agrupamiento de las trayectorias, lo que permite obtener las trayectorias representativas de las agrupaciones, funcionamiento que es también aplicado en la técnica Traclass al realizar el particionado y el agrupamiento basado en trayectorias.

El nivel de precisión de los resultados en Traclass son muchos más efectivos que los brindados en Traclus. Los dos agrupamientos realizados en Traclass deben ser llevados a cabo según el nivel de precisión que se quiera lograr en los resultados, es decir, puede darse el caso de que en diversas circunstancias solamente se necesite realizar el agrupamiento basado en regiones. Para obtener un alto grado de especificidades en las características de las trayectorias se deben realizar los dos agrupamientos debido a que solamente con el agrupamiento basado en regiones es menos probable lograr una alta calidad y eficiencia en la clasificación de las trayectorias.

Al realizar la comparación entre estas dos técnicas se puede evidenciar que la velocidad es mayor en la técnica que más beneficios aporta, sin embargo, esto se debe al mayor nivel de complejidad que posee y



por tanto a la precisión y eficiencia con la que realiza las funcionalidades. En este caso la selección efectiva y de mayor precisión es la técnica Traclass.

### **Técnicas DFS\_MINE y Periódica**

Las técnicas DFS\_MINE y Periódica presentan similitudes en cuanto al objetivo que cumplen al realizar su procedimiento de determinar patrones frecuentes o comportamientos periódicos. Ambas son de gran utilidad en la detección de movimientos futuros y comportamientos inusuales en las entidades móviles. Estas características en aplicaciones para el control de flotas son de gran importancia, además de que constituyen tareas vitales y necesarias para que el proceso de control sea realizado con la mayor eficiencia posible.

A pesar de que estas técnicas funcionen de forma similar de cierta manera para lograr el mismo objetivo, cada una presenta especificidades que se muestran como ventajas. DFS\_MINE es una técnica de gran eficiencia en cuanto a búsqueda y resultados. Esta técnica a pesar de presentar en algunas ocasiones demoras realizando parte de sus algoritmos, es recompensada con respecto a la habilidad que presenta al almacenar la información que necesita, y lograr así trabajar con largas secuencias de patrones frecuentes. La técnica Periódica es rápida, pero posee un método que es sensible al ruido espacial (ver epígrafe 2.3.1), por lo que el nivel de precisión de sus resultados en ocasiones no son los esperados. La eficiencia de esta técnica depende en ocasiones de la información especificada por el usuario o del estudio realizado por la técnica a partir de información almacenada.

Otra de las ventajas que presenta la técnica DFS\_MINE es que, a diferencia de las otras, no consume gran cantidad de recursos de la memoria principal, es decir, que a medida que va aumentando la cantidad de datos a almacenar esta no presenta problemas. Basándose en la comparación realizada se puede llegar a la conclusión de que entre DFS\_MINE y Periódica la selección más efectiva es la técnica DFS\_MINE debido a las ventajas que presentan para el trabajo en aplicaciones para el control de flotas.

### **Técnica DBSCAN**

La técnica DBSCAN posee disímiles ventajas en cuanto al funcionamiento que presenta, logrando así realizar gran cantidad de agrupaciones de forma óptima y efectiva, siendo de esta forma reutilizada en



muchas de las técnicas seleccionadas. Esta técnica dentro de las analizadas en el epígrafe 3.2 es la única que se utiliza para el agrupamiento y por las características que posee descubre satisfactoriamente todos los agrupamientos y detecta los puntos de ruido para una determinada BD a diferencia de otros métodos de agrupamiento.

### **Técnica ObjectGrowth**

Esta técnica presenta disímiles ventajas debido a que en su funcionamiento incluye varias reglas de poda que posibilitan una mayor rapidez y efectividad en la detección de enjambres. Su ventaja se basa en que solamente se encarga de buscar enjambres cerrados, reduciendo el espacio de búsqueda en su gran mayoría. Entre las explicadas en el capítulo tres epígrafe 3.2 la técnica ObjectGrowth es la única que se encarga del reconocimiento de este patrón, siendo de gran utilidad en el trabajo con trayectorias al ser incorporada a los SIG.

### **3.5 Conclusiones Parciales**

En el capítulo se realizó un profundo análisis acerca del funcionamiento de cada técnica caracterizada en el epígrafe 2.2 así como experimentos realizados a dos de estas técnicas lo que evidencia su alta funcionalidad y utilidad en SIG para el control de flotas. El desarrollo del capítulo posee gran importancia ya que se muestra la capacidad que poseen las técnicas de brindar importantes soluciones del trabajo con las trayectorias y su eficiencia en la predicción de movimientos futuros, análisis de recorridos y control de entidades móviles. Los argumentos definidos forman parte de la base complementaria que fue utilizada en la comparación, logrando así que la selección realizada fuese precisa y eficiente.

### CONCLUSIONES GENERALES

Una vez concluida la investigación en la cual se realizó una selección de técnicas de MD espacio-temporal para el análisis de trayectorias de objetos móviles, se puede arribar a las siguientes conclusiones:

- Con la realización de una caracterización detallada de técnicas de minería de datos espacio-temporal se logró una mayor comprensión de su funcionamiento. Esta caracterización posibilitará que la comparación realizada posea argumentos evidentes para realizar una eficaz selección.
- La comparación realizada entre las técnicas posibilitó la obtención de una eficiente y precisa selección de técnicas de minería de datos espacio-temporal que será integrado en un futuro a SIG para el control de flotas.
- El aporte de la investigación para una futura incorporación de técnicas avanzadas de minería de datos espacio-temporal a los SIG lo que permitirá que sean capaces de razonar o analizar datos extraídos de bases de datos, logrando así aportar información acerca de entidades móviles.
- Con el desarrollo de elementos teóricos referentes a la minería de datos espacio-temporal se logró la incorporación de nuevos elementos al cuerpo de conocimiento del tema.

Todos estos resultados posibilitan la incorporación de razonamiento en las aplicaciones para el control de flotas, logrando así, un mejor desarrollo en la capacidad de realizar análisis y procesamiento de información espacio-temporal y el cumplimiento de los objetivos propuestos al principio de la investigación.

### **RECOMENDACIONES**

A modo de recomendación se propone:

- ✓ La utilización de la documentación como material de consulta para la realización de otras investigaciones referentes al tema.
- ✓ La implementación de las técnicas de MD espacio-temporal seleccionadas en la investigación para su futura incorporación a los SIG para el control de flotas logrando que los mismos sean capaces de analizar la trayectoria recorrida por un objeto móvil.

## TRABAJOS REFERENCIADOS

**Aguilera, Manuel Ernesto Acosta. 2009.** *Minería de datos y descubrimiento de conocimiento*. La Habana : Facultad de Economía, Universidad de La Habana, 2009.

**Andrey Tietbohl, Vania Bogorny, Bart Kuijpers, Luis Otavio Álvares. 2008.** *A Clustering-based Approach for Discovering Interesting*. Brasil : s.n., 2008.

**Bartle, Phil. 2007.** Potenciación comunitaria. Supervisión. [En línea] 2007. [Citado el: 16 de 11 de 2010.] <http://www.scn.org/mpfc/modules/mon-miss.htm>.

**Bogorny, Vania. 2010.** Tutorial on Spatial and Spatio-Temporal Data Mining Part II – Trajectory Knowledge Discovery. [En línea] 11 de 10 de 2010. [Citado el: 29 de 1 de 2011.]

**Cerrano Cinca, Carlos. 2010.** Path modeling to bankruptcy: causes and symptoms of the banking crisis. *La contabilidad en la era del conocimiento*. [En línea] 2010. [Citado el: 17 de 11 de 2010.] <http://www.ciberconta.unizar.es/leccion/INTRODUC/485.HTM>.

**Curto, Josep. 2010.** Information Management. [En línea] 28 de 11 de 2010. [Citado el: 16 de 11 de 2010.] <http://informationmanagement.wordpress.com/category/gestion/gestion-de-la-informacion>.

**Gámez, Efraín Ricardo Sosa. 2009.** *Estudio de métodos para proteger y explotar datos de seguimiento de turistas*. Tarragona : Universidad Rovira y Virginia, 2009. Tesis para la obtención del grado de Màster en Ingeniería Informàtica y Seguridad.

**GlobalSat. 2010.** Localización de Flotas GlobalSat. [En línea] 2010. [Citado el: 3 de 12 de 2010.] <http://www.localizacionflotas.com/>.

**Ilias Tsoukatos, Dimitrios Gunopulos. 2001.** *Efficient Mining of Spatiotemporal Patterns*. California : Computer Science Department, University of California Riverside, 2001.

**Jae-Gil Lee, Jiawei Han, Kyu-Young Whang. 2007.** *Trajectory Clustering: A Partition-and-Group Framework*. s.l. : Department of Computer Science University of Illinois at Urbana-Champaign, 2007.

**Jae-Gil Lee, Jiawei Han, Xiaolei Li, Hector González. 2008.** *TraClass: Trajectory Classification Using Hierarchical Region-Based and Trajectory-Based Clustering*. s.l.: Department of Computer Science University of Illinois at Urbana-Champaign, 2008.

**Lázaro, Máx. 2010.** CCIDEP Comité Coordinador de la Infraestructura de Datos Espaciales. [En línea] 2010. [Citado el: 23 de 11 de 2010.] [http://ccidep.gob.pe/index.php?option=com\\_content&view=article&id=18&Itemid=56](http://ccidep.gob.pe/index.php?option=com_content&view=article&id=18&Itemid=56). 15.

**Lic. Osmani Herrera González, Msc. Ing. Rafael Cruz Iglesias, Msc. Ing. José Luis Capote Fernández, Msc. Lic. Guillermo González Suárez, Tec. Liset Becerra Lugones, entre otros. 2008.** MAPPING Revista Internacional de Ciencias de la Tierra. [En línea] 09 de 2008. [Citado el: 3 de 12 de 2010.] [http://www.mappinginteractivo.com/plantilla.asp?id\\_articulo=1516&titulo=&autor=&contenido=&tipo=normal](http://www.mappinginteractivo.com/plantilla.asp?id_articulo=1516&titulo=&autor=&contenido=&tipo=normal) . ISSN: 1131-9100.

**Luis Otavio Alvares, Vania Bogorny, Bart Kuijpers, Jose Antonio Fernandes, Bart Moelans, Alejandro Vaisman. 2006.** A Model for Enriching Trajectories with Semantic Geographical Information. [En línea] 2006. [Citado el: 22 de 02 de 2011.]

**Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu. 1996.** A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. [En línea] 1996. [Citado el: 15 de 03 de 2011.]

**Micronav. 2010.** Micronav Servicio de Control de Flotas mediante GPS/GPRS e Internet. [En línea] 2010. [Citado el: 3 de 12 de 2010.] <http://www.micronav.net/>.

**Song Li, Ph.D. 2009.** *Knowledge Discovery From Trajectories*. 2009. Dissertation work in Geospatial Technologies.

**Stefano Spaccapietra, Christine Parent, Maria Luisa Damiani, Jose Antonio de Macedo, Fabio Porto, Christelle Vangenot. 2007.** *A Conceptual View on Trajectories*. Lausanne, Switzerland : s.n., 2007. 14.

**Tirado, Luis Fernando. 2010.** ArcLogistics™: The Routing and Scheduling Solution for Fleet Management. [En línea] ESRI, 2010. [Citado el: 3 de 12 de 2010.] <http://www.esri.com/library/whitepapers/pdfs/arclogistics-for-fleet.pdf>.

*Una revisión de Bases de Datos con objetos en movimiento.* **Rodríguez, Andrea. 2005.** 11, Universidad de concepción : s.n., 2005.

**Vallejos, Sofia J, Martínez, David Luis la Red. 2006.** *Minería de Datos.* Corrientes - Argentina : s.n., 2006.

**Velazco, Arturo. 2010.** Definición.de. [En línea] 2010. [Citado el: 11 de 11 de 2010.] <http://definicion.de/movil/>.

**Yunta, Luis Rodríguez. 2001.** *Bases de Datos Documentales.* MALDONADO, Angeles : CINDOC, 2001.

**Zhenhui Li, Bolin Ding, Jiawei Han, Roland Kays. 2009.** *Swarm: Mining Relaxed Temporal Moving Object Clusters.* New York : s.n., 2009.

**Zhenhui Li, Bolin Ding, Jiawei Han, Roland Kays, Peter Nye. 2007.** *Mining Periodic Behaviors for Moving Objects.* New York : s.n., 2007.

**BILBIOGRAFÍA CONSULTADA**

**Aguilera, Manuel Ernesto Acosta. 2009.** *Minería de datos y descubrimiento de conocimiento.* La Habana : Facultad de Economía, Universidad de La Habana, 2009.

**Ilias Tsoukatos, Dimitrios Gunopulos. 2001.** *Efficient Mining of Spatiotemporal Patterns.* California : Computer Science Department, University of California Riverside, 2001.

**Jae-Gil Lee, Jiawei Han, Kyu-Young Whang. 2007.** *Trajectory Clustering: A Partition-and-Group Framework.* s.l. : Department of Computer Science University of Illinois at Urbana-Champaign, 2007.

**Jae-Gil Lee, Jiawei Han, Xiaolei Li, Hector González. 2008.** *TraClass: Trajectory Classification Using Hierarchical Region-Based and Trajectory-Based Clustering.* s.l. : Department of Computer Science University of Illinois at Urbana-Champaign, 2008.

**Lázaro, Máx. 2010.** CCIDEP Comité Coordinador de la Infraestructura de Datos Espaciales. [Online] 2010. [Cited: 11 23, 2010.] [http://ccidep.gob.pe/index.php?option=com\\_content&view=article&id=18&Itemid=56](http://ccidep.gob.pe/index.php?option=com_content&view=article&id=18&Itemid=56). 15.

**Lic. Osmani Herrera González, Msc. Ing. Rafael Cruz Iglesias, Msc. Ing. José Luis Capote Fernández, Msc. Lic. Guillermo González Suárez, Tec. Liset Becerra Lugones, entre otros. 2008.** MAPPING Revista Internacional de Ciencias de la Tierra. [Online] 09 2008. [Cited: 12 3, 2010.] [http://www.mappinginteractivo.com/plantilla.asp?id\\_articulo=1516&titulo=&autor=&contenido=&tipo=normal](http://www.mappinginteractivo.com/plantilla.asp?id_articulo=1516&titulo=&autor=&contenido=&tipo=normal) . ISSN: 1131-9100.

**Luis Otavio Alvares, Vania Bogorny, Bart Kuijpers, Jose Antonio Fernandes, Bart Moelans, Alejandro Vaisman. 2006.** A Model for Enriching Trajectories with Semantic Geographical Information. [Online] 2006. [Cited: 02 22, 2011.]

**Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu. 1996.** A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. [Online] 1996. [Cited: 03 15, 2011.]

**Micronav. 2010.** Micronav Servicio de Control de Flotas mediante GPS/GPRS e Internet. [Online] 2010. [Cited: 12 3, 2010.] <http://www.micronav.net/>.

**Song Li, Ph.D. 2009.** *Knowledge Discovery From Trajectories*. 2009. Dissertation work in Geospatial Technologies.

**Stefano Spaccapietra, Christine Parent, Maria Luisa Damiani, Jose Antonio de Macedo, Fabio Porto, Christelle Vangenot. 2007.** *A Conceptual View on Trajectories*. Lausanne, Switzerland : s.n., 2007. 14.

**Tirado, Luis Fernando. 2010.** ArcLogistics™: The Routing and Scheduling Solution for Fleet Management. [Online] ESRI, 2010. [Cited: 12 3, 2010.] <http://www.esri.com/library/whitepapers/pdfs/arclogistics-for-fleet.pdf>.

*Una revisión de Bases de Datos con objetos en movimiento.* **Rodríguez, Andrea. 2005.** 11, Universidad de concepción : s.n., 2005.

**Vallejos, Sofia J, Martínez, David Luis la Red. 2006.** *Minería de Datos*. Corrientes - Argentina : s.n., 2006.

**Velazco, Arturo. 2010.** Definición.de. [Online] 2010. [Cited: 11 11, 2010.] <http://definicion.de/movil/>.

**Zhenhui Li, Bolin Ding, Jiawei Han, Roland Kays. 2009.** *Swarm: Mining Relaxed Temporal Moving Object Clusters*. New York : s.n., 2009.

**Zhenhui Li, Bolin Ding, Jiawei Han, Roland Kays, Peter Nye. 2007.** *Mining Periodic Behaviors for Moving Objects*. New York : s.n., 2007.



## GLOSARIO DE TÉRMINOS

**Base de datos (BD):** Conjunto de datos pertenecientes que son almacenados sistemáticamente para su posterior utilización.

**CB-SMOT:** Técnica para descubrir lugares interesantes en las trayectorias a partir de agrupaciones.

**DBSCAN:** Técnica basada en la densidad para el descubrimiento de grupos en grandes base de datos espaciales con ruido.

**DFS\_MINE:** Técnica utilizada para el descubrimiento efectivo de patrones espacio-temporales.

**GEySED:** Centro de la facultad número seis compuesto por dos departamentos: Señales Digitales y Geoinformática.

**Global Positioning System (GPS):** Sistema de posicionamiento global que permite localizar la ubicación de un objeto en cualquier parte del mundo.

**KDD:** Proceso de descubrimiento de conocimientos en bases de datos.

**Minería de Datos (MD):** Proceso de extracción de información que reside de manera implícita en los datos.

**Objectgrowth:** Técnica para el descubrimiento de grupos de objetos en movimiento que se trasladan temporalmente.

**Roi:** Regiones de interés.

**Sistema de Información Geográfica (SIG):** Integración de hardware y software diseñado para capturar, almacenar, y desplegar información geográfica.

**SMOT:** Técnica para el enriquecimiento de las trayectorias con la semántica de la información geográfica.

**Traclass:** Técnica para realizar clasificación de las trayectorias utilizando métodos de agrupamiento basado en regiones y en trayectorias.

**Traclus:** Técnica de agrupamiento de trayectorias.