

LUPA ESPACIO-TEMPORAL: UNA HERRAMIENTA PARA EL ANÁLISIS VISUAL DE TRAYECTORIAS EN UNA BODEGA DE DATOS

A SPATIO-TEMPORAL MAGNIFYING GLASS: A TOOL FOR THE VISUAL ANALYSIS OF TRAJECTORIES IN A DATAWAREHOUSE

Francisco Javier Moreno¹, Sergio Bota Sierra², Andrés Felipe Cadavid Agudelo³

Fecha de recepción: 19 de junio de 2014

Fecha de aprobación: 12 de octubre de 2014

Referencia: F.J. Moreno, S. Bota Sierra, A.F. Cadavid Agudelo. (2014). Lupa espacio-temporal: una herramienta para el análisis visual de trayectorias en una bodega de datos. *Ciencia e Ingeniería Neogranadina*, 24 (2), pp. 77-96

RESUMEN

En este artículo se presenta una herramienta visual para facilitar el análisis de trayectorias y el descubrimiento de patrones espacio-temporales a partir de una bodega de datos de trayectorias (BDT). La herramienta propuesta, una lupa espacio-temporal, permite que el analista se enfoque en una determinada región donde han ocurrido varias trayectorias y permite detectar, según ciertos parámetros especificados por el analista a través de una interfaz gráfica, p. ej. la relación de cercanía de una trayectoria con otras o con los sitios a su alrededor. En el artículo se proponen y definen formalmente las relaciones de cercanía derivadas entre trayectorias y entre trayectorias y sitios. Una contribución adicional fue el enriquecimiento de un modelo de una BDT con el fin de permitir la formulación de consultas más expresivas y apoyar el aspecto de visualización de la herramienta propuesta. Aunque se requieren experimentos más exhaustivos, los resultados evidenciaron algunos patrones espacio-temporales que demuestran la conveniencia y la utilidad de la herramienta.

Palabras Clave: trayectorias, análisis de trayectorias, episodios, episodios de estadía, episodios de cercanía, bodegas de datos, visualización, espacio-temporal.

-
1. Ingeniero de Sistemas, Ph.D. en Ingeniería, Profesor Asociado. Universidad Nacional de Colombia Sede Medellín, Medellín, Colombia, fjmoreno@unal.edu.co
 2. Ingeniero de Sistemas. Universidad Nacional de Colombia Sede Medellín, Medellín, Colombia, sbotas@unal.edu.co
 3. Ingeniero de Sistemas. Universidad Nacional de Colombia Sede Medellín, Medellín, Colombia, anfcadavidag@unal.edu.co

ABSTRACT

This paper presents a visual tool to facilitate the trajectory analysis and the discovery of spatio-temporal patterns in a trajectory data warehouse (TDW). The proposed tool is a spatio-temporal magnifying glass that allows analysts to focus on a specific region, where several trajectories have occurred and to detect, according to some parameters specified by the analyst through a graphical interface, e.g. the closeness relationship between trajectories or between a trajectory and its surrounding sites. In this paper, we propose and formally define derived closeness relationships between trajectories and between trajectories and sites. An additional contribution was the enrichment of a TDW model in order to allow the formulation of more expressive queries and to support the visualization aspect of the proposed tool. Although experiments that are more exhaustive are required, our results evidenced some spatio-temporal patterns that demonstrate the convenience and advantages of our tool.

Keywords: trajectories, trajectory analysis, episodes, stop episodes, proximity episodes, data warehouses, visualization, spatio-temporal

INTRODUCCIÓN

En la última década, las bodegas de datos (BD) han demostrado su utilidad para la integración y análisis de grandes volúmenes de datos [1]. Por otro lado, gracias a los equipos móviles con dispositivos de rastreo [2] y sistemas de información geográficos [3], actualmente es posible recopilar grandes volúmenes de datos sobre el movimiento de objetos móviles como turistas, animales, vehículos, entre otros. El movimiento de un objeto móvil se puede describir mediante trayectorias. Según Spaccapietra [4], una trayectoria es el registro de la evolución de la posición de un objeto que se mueve en un espacio durante un periodo con el fin de lograr un objetivo. Por ejemplo, un turista que se mueve en una ciudad en la mañana con el fin de visitar determinados sitios turísticos. Esta definición guarda una estrecha relación con el concepto de trayectoria (*path*) de la geografía temporal (*time geography*) [5], en el que "en el espacio-tiempo los individuos describen una trayectoria" en un contexto

situacional y están sujetos a un conjunto de restricciones (físicas, sociales, impuestas por las autoridades, etc.).

Una trayectoria se puede segmentar a su vez en episodios. Por ejemplo, si una trayectoria ingresa a un sitio s y permanece allí (continuamente) durante un mínimo de tiempo Δt entonces se dice que la trayectoria tuvo un episodio de estadía (*stop*) en s (p. ej. un vehículo que ingresa a una gasolinera y permanece en esta mínimo 5 min). Por otro lado, si una trayectoria permanece (continuamente) a menos de una distancia Δd de un sitio s durante un mínimo de tiempo Δt (sin ingresar a s durante Δt) entonces se dice que la trayectoria tuvo un episodio de cercanía (*proximity*) con respecto a s (p. ej. un vehículo que permanece fuera de un restaurante, a menos de 50 m, durante mínimo 5 min).

Considérese, p. ej. las trayectorias de los vehículos durante un día en una ciudad. Supóngase que se desea saber p. ej. cuántas

y cuáles de esas trayectorias: a) ingresaron a una gasolinera y permanecieron allí mínimo 5 min, i. e. tuvieron un episodio de estadía en una gasolinera y b) permanecieron a menos de 50 m de un restaurante durante mínimo 3 min (tiempo durante el cual no ingresaron al restaurante), i. e. tuvieron un episodio de cercanía con respecto a un restaurante. La visualización de los resultados de estas consultas podría ayudar a identificar patrones espacio-temporales. Por ejemplo, analizar el comportamiento de las trayectorias de los vehículos en el centro de una ciudad en días laborales. El análisis puede ser con respecto a otras trayectorias o a eventos (manifestaciones, congestiones vehiculares) o a sitios que se encuentran en el espacio donde las trayectorias en análisis ocurren.

Las trayectorias se han incorporado como elementos de análisis en las BD, dando origen a las BD de trayectorias (BDT) [6]. Las BDT exigen el desarrollo de estructuras de datos para el almacenamiento y mecanismos de indexación eficientes de las trayectorias [7], modelos conceptuales [6,8], operadores OLAP [9-10] y de agregación [11] y mecanismos de visualización [12] (ver también la subsección 4.2) que faciliten la comprensión y la identificación de posibles tendencias y patrones espacio-temporales [13] (ver también el proyecto MOVE, <http://move-cost.info>, concebido para extraer conocimiento a partir de grandes volúmenes de datos sobre objetos móviles). El presente artículo se enfoca en el aspecto de visualización de trayectorias de una BDT. Este aspecto puede ayudar a comunicar y a descubrir relaciones espacio-temporales que no son percibidas ni están almacenadas explícitamente en la BDT. Este descubrimiento de conocimiento, a partir de despliegues visuales aunado con el conocimiento de los analistas puede ayudar

al planteamiento de hipótesis que expliquen o predigan un determinado comportamiento.

Por ejemplo, si se considera que solo a mediodía (en los días laborales) en Nueva York hay cerca de 220 000 vehículos en circulación [14] y en Londres 155 000 [15] estos volúmenes de datos exigen mecanismos de visualización que faciliten su análisis, p. ej. en términos de congestión. En este artículo, se propone una herramienta de visualización, una lupa espacio-temporal, que puede ayudar a los analistas en este tipo de tareas.

La lupa propuesta permite que el analista se enfoque en una región arbitraria del espacio donde ocurrieron las trayectorias y detalle los resultados de una determinada consulta previamente parametrizada. Por ejemplo, considérese la consulta "mostrar todas las trayectorias que ocurrieron en una ciudad durante una semana". Si los resultados muestran, p. ej. que en el centro de la ciudad confluyen más trayectorias que en los barrios periféricos, el analista puede delimitar y enfocar con la lupa la región correspondiente al centro de la ciudad y así obtener más detalles sobre el comportamiento de las trayectorias en esa región de la ciudad. Esta característica (zoom, aumento) también está presente en los SIG y en la geovisualización (un área de desarrollo reciente de la geografía temporal y que puede ayudar en la exploración de datos y en los procesos de toma de decisiones [16]). La lupa también permite restringir los resultados temporalmente. En el ejemplo anterior, el analista puede enfocar el centro de la ciudad e indicar que solo desea obtener resultados para un día específico de la semana analizada.

Además, en el artículo se proponen las relaciones de cercanía derivadas para ayudar en el análisis de trayectorias. Este tipo de relaciones se dice

que son derivadas porque no están almacenadas explícitamente en la BDT, i. e. son relaciones que pueden ser descubiertas a partir de consultas. En este sentido, la propuesta se relaciona con la generación de conocimiento a partir de conocimiento almacenado, uno de los aspectos clave de la geovisualización [16]. En este trabajo se consideran dos tipos de relaciones de cercanía derivadas: i) entre un episodio de una trayectoria y un sitio, p. ej. sea la consulta "mostrar los episodios de estadía que duraron mínimo 5 min en una gasolinera que está a menos de 500 m de un parque"; y ii) entre episodios, p. ej. sea la consulta "mostrar los episodios de estadía que duraron mínimo 5 min en una gasolinera que está a menos de 500 m de un parque donde ocurrieron episodios de estadía que duraron mínimo 5 min el mismo día". Así, las relaciones de cercanía derivadas pueden ayudar a descubrir patrones de comportamiento espacio-temporal basados en proximidad.

El resto del artículo está organizado así. En la Sección 1 se presenta un ejemplo motivador donde se expone y enriquece un modelo para la gestión de trayectorias en una BDT. En la Sección 2 se presentan los elementos de la lupa espacio-temporal y se definen las relaciones de cercanía derivadas. En la Sección 3 se presentan algunos experimentos y su respectivo análisis. En la Sección 4 se presenta el estado del arte. Finalmente, en la Sección 5 se concluye el artículo y se plantean trabajos futuros.

1. EJEMPLO MOTIVADOR

En esta sección se presenta un modelo para una BDT propuesto por Campora et al. [17]. En el presente artículo, dicho modelo fue enriquecido con elementos que permiten el planteamiento de consultas como las expuestas en la subsección 1.3. Las mejoras específicas (ya

incorporadas al modelo de la subsección 1.1.) se detallan en la subsección 1.2.

1.1. MODELO CONCEPTUAL PARA UNA BDT

Considérese el modelo de la Figura 1 propuesto por Campora [17]. El modelo usa la notación de Malinowski [18] para la representación de BD espacio-temporales. El modelo se enfoca en el diseño e implementación de una BDT para el análisis del tráfico de vehículos en una ciudad, aunque en general se puede aplicar para cualquier objeto móvil.

Para el caso particular de una trayectoria, un episodio es una secuencia de puntos (x, y, t) que presentan un comportamiento común según un criterio establecido por el analista, p. ej. el objeto permanece continuamente dentro de un sitio durante un tiempo mínimo dado (i. e. un episodio de estadía). Además un episodio se relaciona con las siguientes dimensiones:

- a) Evento o sitio: permite analizar los episodios con respecto a eventos tales como una manifestación, un concierto, una congestión vehicular o a sitios como un restaurante, un parque, entre otros. Por ejemplo, analizar si un episodio ocurrió dentro o cerca de una manifestación. Esto podría explicar un promedio de velocidad inusualmente bajo por parte de un vehículo en una determinada zona. Los eventos y sitios se clasifican en categorías, p. ej. eventos o sitios de entretenimiento, políticos, entre otros.
- b) Trayectoria: indica la trayectoria a la que pertenece el episodio. Las trayectorias se clasifican en grupos. Cada trayectoria es hecha por un objeto móvil.
- c) Geográfica: indica dónde ocurrió un episodio. Un episodio ocurre en una división espacial, p. ej. una ciudad, la cual puede estar contenida en otra división espacial, p.

ej. un estado, y esta a su vez puede estar contenida en otra división espacial, p. ej. un país.

- d) Temporal: indica cuándo ocurrió un episodio. Un episodio ocurre en un día. Los días se agrupan en meses y los meses en años. Dependiendo de las necesidades de análisis se podría definir un nivel de tiempo menor, p. ej. hora, minuto, segundo, etc.

un lenguaje visual) cómo estos fenómenos ocurren en un marco espacio-temporal sujetos a un conjunto de restricciones.

En la Tabla 1 y en la Figura 2 se muestra un ejemplo de episodios (hechos) correspondientes al modelo de la Figura 1.

1.2. MEJORAS INCORPORADAS AL MODELO DE CAMPORA

Aunque aquí se distingue entre trayectorias, episodios y eventos, desde un punto de vista más general cada uno de estos conceptos se refiere a un fenómeno que ocurre en el espacio-tiempo. Así, nuestra propuesta ofrece un mecanismo para analizar posibles relaciones entre estos fenómenos de manera similar a como en la geografía temporal (donde una trayectoria se segmenta en eventos y desplazamientos) se estudia (mediante

El modelo de Campora fue enriquecido con los siguientes elementos (ver Figura 1):

- Se agregó el nivel Event/Site Category para clasificar los eventos y sitios. Se incluyó un atributo Icon para representar gráficamente la categoría del evento o sitio (manifestación, restaurante, parque, entre otros).
- Se agregaron los atributos Min. Stop Time (i. e. tiempo mínimo de estadía), Min. Closeness Time y Min. Closeness Distance

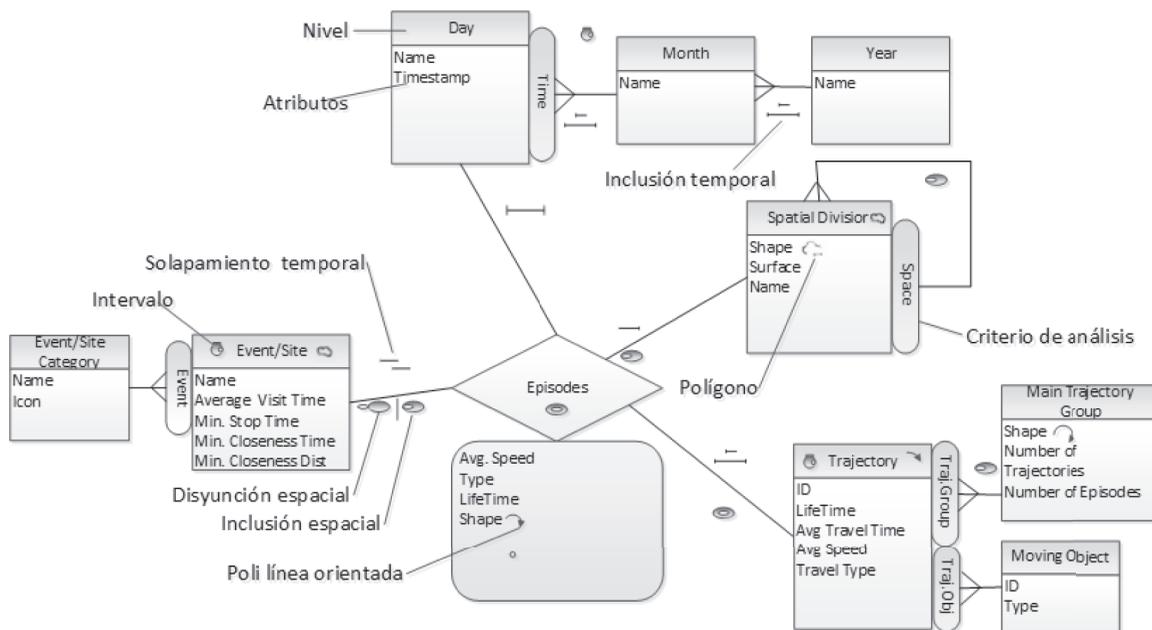


Figura 1. Modelo para una BDT.
Fuente: Elaboración propia (modificado de [17]).

Tabla 1. Muestra de episodios de la BDT de la Figura 1.

Hecho	Medidas				Dimensiones			
Episodio	Avg. Speed	Type	LifeTime	Shape	Event	Trajectory	Spatial division	Day
...
10	2 km/h	Cercanía	9 min	→	Rte. Azul	28	NY City	26/Ago/2013
11	0.01 km/h	Estadía	40 min	●	Rte. Azul	28	NY City	26/Ago/2013
12	2 km/h	Cercanía	9 min	→	Rte. Azul	28	NY City	26/Ago/2013
13	2.5 km/h	Cercanía	5 min	↗	Parque Central	28	NY City	26/Ago/2013
14	3 km/h	Cercanía	3 min	→	Parque Chino	28	NY City	26/Ago/2013
15	0 km/h	Estadía	2 h	●	Parque Chino	28	NY City	26/Ago/2013
...
29	0.01 km/h	Estadía	1 h	●	Parque Central	32	NY City	26/Ago/2013

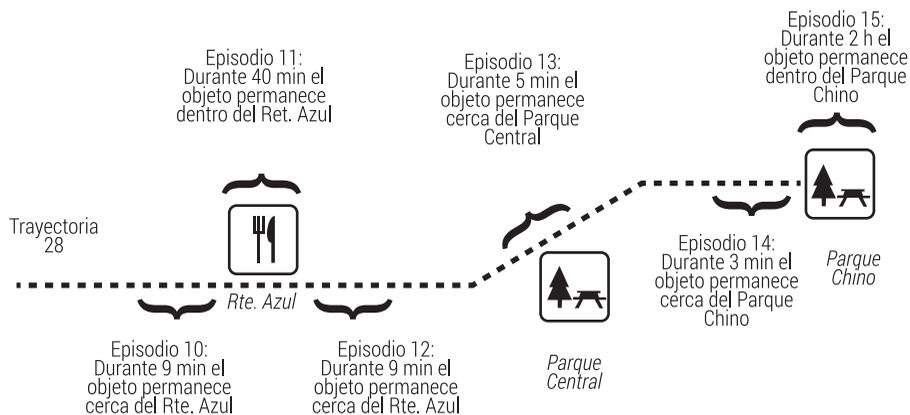


Figura 2. Episodios de la trayectoria 28 (línea punteada) correspondientes a la Tabla 1.

al nivel Event/Site.

- Se agregó el nivel Moving Object. El atributo Type indica si el objeto es una persona, un animal, un vehículo, etc.
- La relación entre los episodios y los eventos o sitios permite tanto inclusión como disyunción espacial (el modelo original solo consideraba esta última). Esto permite que un episodio se relacione no solo con eventos/sitios que ocurrieron/están cerca sino también con eventos o sitios donde el episodio ocurrió adentro, p. ej. cuando el objeto móvil ingresa a un restaurante.
- En el modelo original, los episodios están relacionados con una dimensión espacial con tres niveles (divisiones espaciales fijas) Area, Region y Environment, en los que una región se compone de áreas y un ambiente de regiones.

Como estos niveles son estructuralmente isomorfos, se definió un solo nivel Spatial Division con una relación recursiva. Esto simplifica el modelo y permite que a) haya cualquier número de niveles en la jerarquía espacial y b) un episodio se relacione con cualquier división espacial y no solo con una división ubicada en el nivel más inferior de la jerarquía espacial.

1.3. CONSULTAS ANALÍTICAS SOBRE LA BDT

La BDT cuyo modelo conceptual se presentó en la subsección 1.1, permite el planteamiento de consultas como:

Consulta 1: “Mostrar todos los episodios de tipo estadía que a) duraron más de 3 min, b) ocurrieron a menos de 50 m de sitios de tipo ‘Restaurante’ y c) ocurrieron dentro de 3 km del Duomo en Milán”.

Debido al gran número resultados que la consulta puede generar y para facilitar la visualización, Campora opta por una muestra de los resultados

(1000 trayectorias). Aunque trabajar con una muestra significativa de datos puede ser suficiente para algunas aplicaciones, también se puede perder información valiosa [19]. Además, aun con una muestra de datos, en su propuesta persisten los problemas de visualización, p. ej. para el usuario no es claro cuáles son los restaurantes que están cerca de cada episodio de tipo estadía (puntos amarillos en la Figura 3).

Ahora supóngase que el usuario desea:

- Detallar los resultados de la Consulta 1. Por ejemplo, dado un episodio de tipo estadía que ocurrió en un restaurante específico, conocer las características de éste: nombre, tiempo promedio de visita, categoría, entre otras.
- Conocer durante un periodo la distribución de los episodios de tipo estadía, p. ej. obtener para cada día de una semana específica el número total de episodios de tipo estadía que ocurrieron en cada restaurante.
- Detallar (drill down espacial) una región en particular (p. ej. delimitar una región en la Figura 3) y obtener información como



Figura 3. Resultados de la Consulta 1.

Fuente: [17].

el número y el tiempo promedio de los episodios que ocurrieron en esa región.

Otras consultas que pueden ser útiles y cuya visualización de resultados juega un papel clave para facilitar el análisis son:

Consulta 2: "Mostrar todos los episodios de tipo estadía que a) ocurrieron en un día específico (p. ej. el lunes 26 de agosto de 2013), b) duraron más de 5 min y c) ocurrieron a menos de 500 m de sitios de tipo 'Restaurante', 'Parque' y 'Manifestación'".

Consulta 3: "Mostrar todos los episodios de tipo estadía que a) ocurrieron el lunes 26 de agosto de 2013, b) duraron más de 5 min y c) ocurrieron a menos de 500 m de episodios de tipo estadía de otras trayectorias".

Para facilitar lo anterior se propone una lupa espacio-temporal.

2. HERRAMIENTA PROPUESTA: LUPA ESPACIO-TEMPORAL

La herramienta propuesta se basa en el concepto de lupa, que esencialmente es un instrumento óptico que permite que un objeto se vea más grande y que se puedan apreciar más detalles del mismo. Siguiendo con este principio, se propone una herramienta para facilitar la visualización de los resultados de consultas espacio-temporales de la BDT presentada. La lupa se puede aplicar a los resultados de una consulta: el usuario se puede enfocar (delimitar) en una región arbitraria y observar con mayor detalle los datos relacionados con la consulta (drill down espacial).

Las características de la lupa son:

- Marco: permite delimitar la región de acción

de la lupa. Actualmente, el tamaño del marco es fijo. Sin embargo, una de las mejoras planeadas es que pueda ser modificada por el analista.

- Mango: permite desplazar y enfocar la lupa por el ambiente geográfico (dimensión espacial) del modelo.
- Atributos: los atributos e indicadores de la lupa se corresponden con los atributos del modelo presentado y se incluyen de manera predeterminada. Los atributos incluidos son el número total de episodios contenidos en el marco de la lupa, el número total de trayectorias, el número total de eventos o sitios, los promedios de duración y velocidad de los episodios, el promedio de visitas a los eventos o sitios, así como información clasificada por tipos de episodio, p. ej. el número de episodios de tipo estadía.
- Formas geométricas: para distinguir las categorías de los eventos o sitios se emplean los iconos incorporados al modelo en el nivel Event/Site Category.
- Interfaz: es un panel auxiliar donde el usuario puede visualizar información de los atributos enmarcados por la lupa.
- Temporal: El usuario puede desplazar la barra temporal ubicada en la interfaz, para ver cómo cambia el valor de los atributos a través del tiempo. La mínima unidad de desplazamiento temporal corresponde al nivel inferior de la dimensión tiempo (día, en el ejemplo).

Ejemplo 1. La lupa aplicada sobre los resultados de la Consulta 1, ver Figura 4, muestra para cada episodio de tipo estadía sus restaurantes cercanos. Además, se muestran medidas adicionales (ver paneles a la derecha) que pueden ser útiles para la toma de decisiones: el número total de trayectorias, el número total de eventos, el número total de episodios contenidos en la región enfocada, los promedios de velocidad y de duración de los episodios.

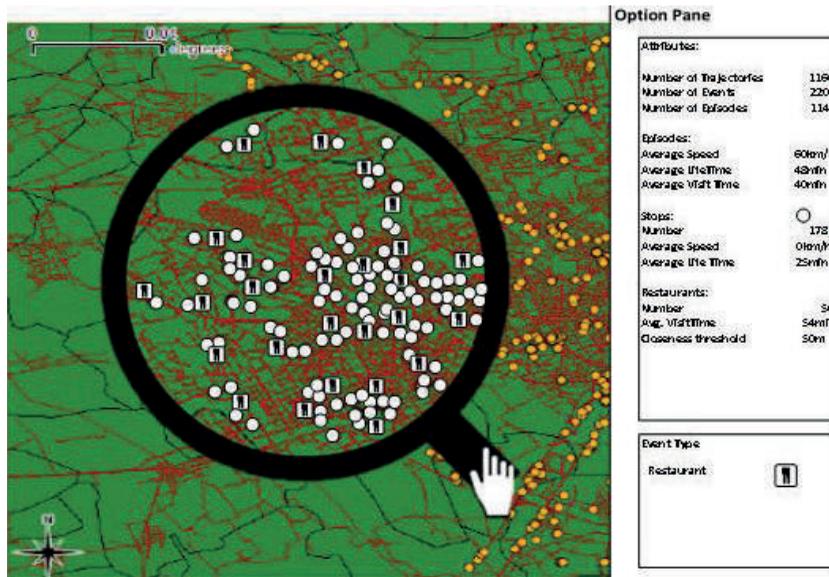


Figura 4. Aplicación de la lupa a los resultados de la Consulta 1.

Ejemplo 2. Supóngase que la lupa se aplica sobre los resultados de la Consulta 2.

La lupa, ver Figura 5, muestra para cada episodio de tipo estadía (que cumplió la fecha y duración especificadas) sus eventos cercanos (que están a menos de 500 m). Además, se muestran datos como el número total de episodios en la región enfocada, el número total de trayectorias, el número total de eventos, entre otros. El analista también puede seleccionar otra fecha, p. ej. el domingo 25 de agosto de 2013, ver Figura 6, y así puede comparar los resultados entre ambas fechas (un día laboral vs. un día festivo).

Nótese que la Consulta 1 y la Consulta 2 pueden revelar p. ej. relaciones de cercanía que no están almacenadas en la BDT. Por ejemplo, considérese de nuevo los episodios de la trayectoria 28 de la Tabla 1 y de la Figura 2. Supóngase que esta trayectoria ocurrió

el 26 de agosto de 2013. Por lo tanto, los episodios 11 y 15 satisfacen las condiciones a) y b) de la Consulta 2. Supóngase además que el Parque Central se encuentra a 400 m del Parque Chino y a 600 m del Restaurante Azul. De esta forma solo el episodio 15 (estadía en el Parque Chino) satisface la condición c) de la consulta, es decir, según los parámetros de la consulta, se descubre una relación de cercanía derivada (que no está almacenada en la BDT) a partir de la cercanía de un episodio de tipo estadía de la trayectoria con respecto a un evento o sitio y a un umbral de proximidad especificado por el analista en la consulta, véase la Figura 7.

También se pueden descubrir relaciones de cercanía entre episodios de trayectorias. Por ejemplo, considérese la Consulta 3. En la Figura 8 se muestran dos trayectorias que tienen episodios de tipo estadía que satisfacen las condiciones de la Consulta 3.

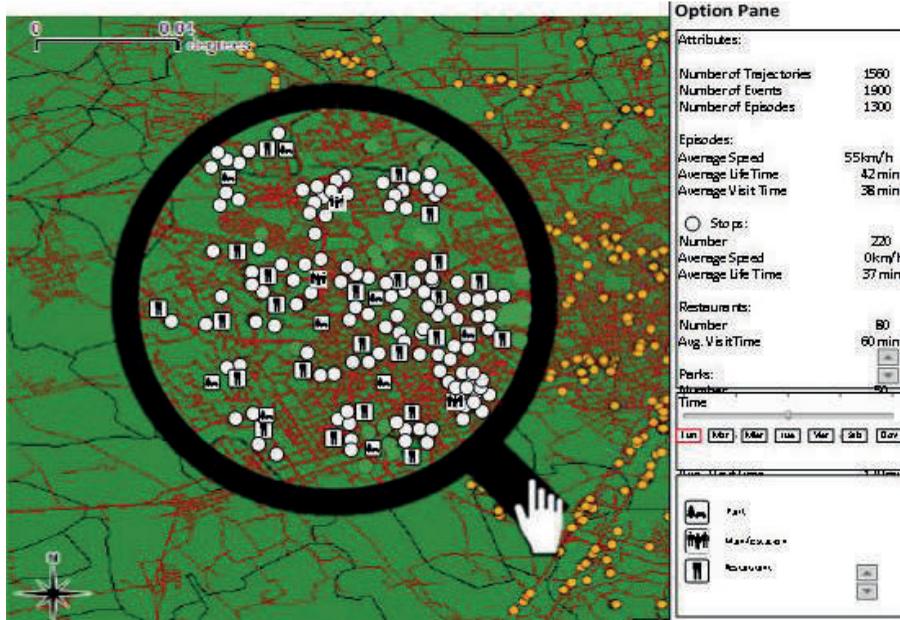


Figura 5. Aplicación de la lupa a los resultados de la Consulta 2 (para el lunes 26/Ago/2013).

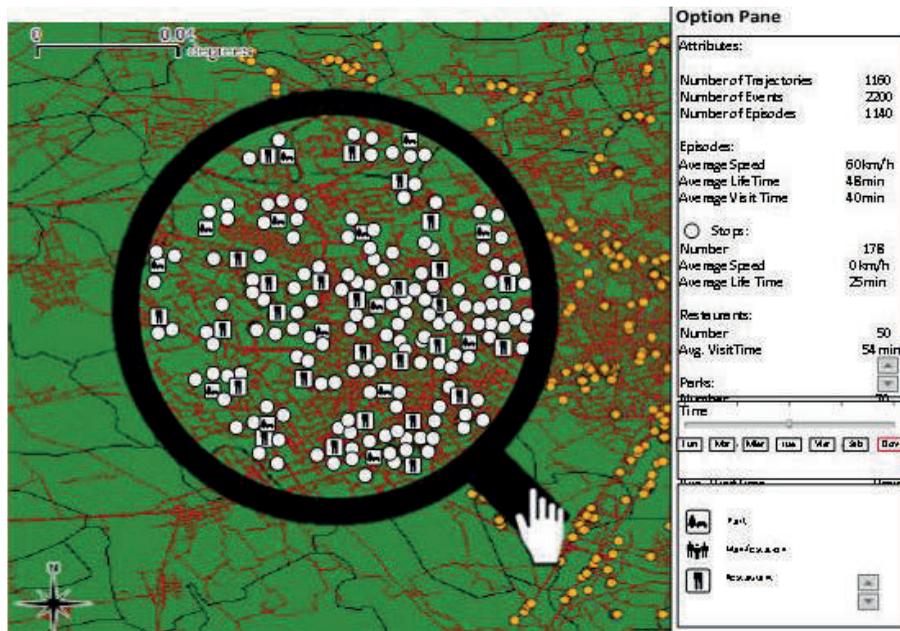


Figura 6. Aplicación de la lupa a los resultados de la Consulta 2 (para el domingo 25/Ago/2013).

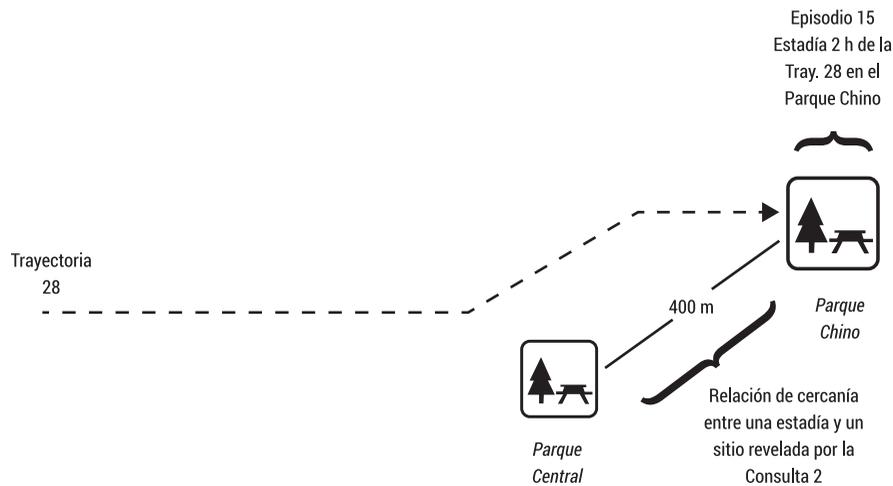


Figura 7. Resultado de la Consulta 2 para la Trayectoria 28.

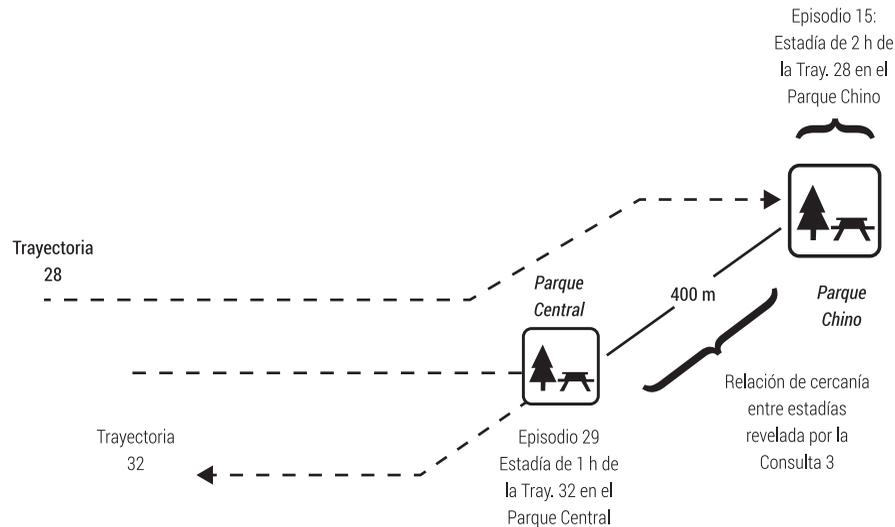


Figura 8. Resultado de la Consulta 3 para las trayectorias 28 y 32.

De esta forma, el analista puede descubrir relaciones de cercanía entre los episodios de una trayectoria $t1$ con respecto a eventos o sitios o a los episodios de una trayectoria $t2$ ($t1$ y $t2$ no necesariamente distintas).

Formalmente, una relación de cercanía derivada entre un episodio de una trayectoria

y un evento o sitio es una tupla $(ep, evs, \Delta dist)$ donde ep es un episodio de una trayectoria, evs es un evento o sitio y $\Delta dist$ es el umbral de cercanía entre ep y evs , ver Figura 9a).

Ejemplo 3. Considérese la Consulta 2 y la Figura 7, entonces (Episodio 15, Parque Central, 500 m) es una relación de cercanía derivada entre

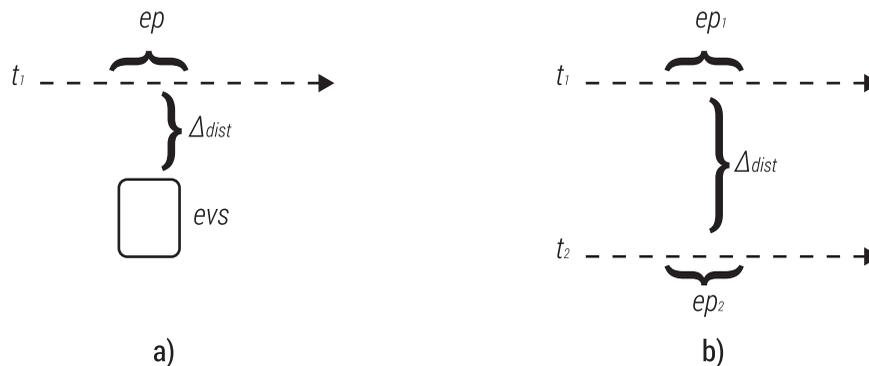


Figura 9. Relaciones de cercanía derivadas: a) entre un episodio y un evento y b) entre episodios.

un episodio de una trayectoria y un evento o sitio.

Formalmente, una relación de cercanía derivada entre un episodio de una trayectoria $t1$ y un episodio de una trayectoria $t2$ ($t1$ y $t2$ no necesariamente distintas) es una tupla $(ep1, ep2, \Delta dist)$ donde $ep1$ es un episodio de $t1$, $ep2$ es un episodio de $t2$ y $\Delta dist$ es el umbral de cercanía entre $ep1$ y $ep2$, ver Figura 9 b). La distancia entre $ep1$ y $ep2$ corresponde a la distancia entre los eventos o sitios asociados con $ep1$ y con $ep2$, respectivamente.

Ejemplo 4. Considérese la Consulta 3 y la Figura 8, entonces (Episodio 15, Episodio 29, 500 m) es una relación de cercanía derivada entre episodios de dos trayectorias.

Nótese que la generación de relaciones de cercanía depende de los parámetros especificados por el analista. Por ejemplo, el analista podría indicar, aparte del umbral de cercanía ($\Delta dist$), parámetros como la fecha y lugar de ocurrencia de los episodios, su tipo, su tiempo mínimo de duración, entre otros. La herramienta propuesta, facilita tanto la especificación de estos parámetros, ver Figura 10, como la visualización de los resultados.

3. EXPERIMENTOS

Un primer experimento se hizo en la ciudad de Medellín. Se seleccionaron 21 sitios de interés (entre restaurantes y hoteles) y con la ayuda de taxistas de la ciudad se recopilieron datos sobre las visitas de pasajeros a estos sitios. El número de trayectorias (cada carrera corresponde a una trayectoria) fue 73: 40 un domingo y 33 un lunes. Para la identificación de episodios de tipo estadía se consideró para todos los sitios Min. Stop Time = 5 min y para los episodios de tipo cercanía Min. Closeness Time = 30 seg y Min. Closeness Dist = 20 m.

Se obtuvieron 416 episodios en total, de los cuales 96 fueron episodios de estadía: 55 el domingo y 41 el lunes (usualmente, el pasajero solo ingresaba a uno de los sitios de interés que era de hecho el destino de la carrera, y solo en algunas ocasiones el pasajero durante la carrera ingresaba a otros de los sitios de interés seleccionados). Por otro lado, hubo 320 episodios de cercanía: 176 el domingo y 144 el lunes. Esto es razonable porque i) cada aproximación del taxi hacia el sitio de destino de la carrera generaba efectivamente una cercanía con respecto a dicho sitio y ii) durante una

Figura 10. Interfaz de ingreso de parámetros para las consultas de relaciones de cercanía derivadas.

carrera el taxi pasaba cerca (a menos de 20 m) de varios de los sitios de interés seleccionados (y se mantenía cerca de estos durante mínimo 30 seg). Por lo tanto, los datos indican un promedio de 4,4 cercanías por carrera el domingo y 4,36 el lunes. Nótese que si el Δt para los episodios de tipo cercanía hubiese sido mayor a 30 seg, p. ej. 3 min, el número de estos episodios hubiese sido mucho menor ya que un taxi durante su desplazamiento difícilmente se mantiene cerca de un sitio por más de 3 min (a menos que hubiese, p. ej. una congestión de tráfico cerca del sitio por donde el taxi se desplaza).

El siguiente paso fue construir un prototipo para experimentar con los datos recogidos. La BDT se implementó en el SGBD PostgreSQL, para los aspectos de visualización se usó QGIS y GoogleMaps, y las interfaces se desarrollaron en Java. El prototipo se puede descargar desde <https://www.dropbox.com/s/pxv7h1kjoomp8bn/Lupa.zip>.

Al formular la Consulta 2 para el lunes se obtuvieron los resultados mostrados en la Tabla 2 y en la Figura 11. El usuario puede seleccionar mediante el mouse un episodio y el aplicativo muestra las estadías cercanas (el usuario debe ingresar un umbral de distancia) con respecto a la estadía seleccionada, i. e. estas corresponden a relaciones de proximidad derivadas entre episodios de trayectorias. El prototipo también permite que el usuario seleccione un episodio y se muestran los sitios cercanos, i. e. relaciones de proximidad derivadas entre un episodio y un sitio de interés de los seleccionados.

La Tabla 2 muestra algunos resultados consolidados para los dos días. Aunque se requieren experimentos con más datos, se pudo observar, p. ej. que el tiempo promedio de las carreras fue mayor el lunes que el domingo, indicador a), esto se podría explicar porque algunas carreras ocurrieron en horas

Tabla 2. Resultados consolidados del experimento 1.

Indicador	Domingo	Lunes
a) Tiempo promedio por trayectoria (carrera)	33 min	41 min
b) Tiempo promedio de las cercanías	52 seg	88 seg
c) Número promedio de estadías (por trayectoria) que ocurrieron cerca de otra estadía considerando un radio de 100 m	1,2	1,19
d) Número promedio de sitios de interés cercanos a una estadía considerando un radio de 100 m	1,4	1,42
e) Considerando solo las trayectorias cuyo destino final fue un hotel, ¿cuál fue el número promedio de cercanías (por trayectoria) a hoteles?	2,1	2
f) Considerando solo las trayectorias cuyo destino final fue un restaurante, ¿cuál fue el número promedio de cercanías (por trayectoria) a restaurantes?	2,7	2,4

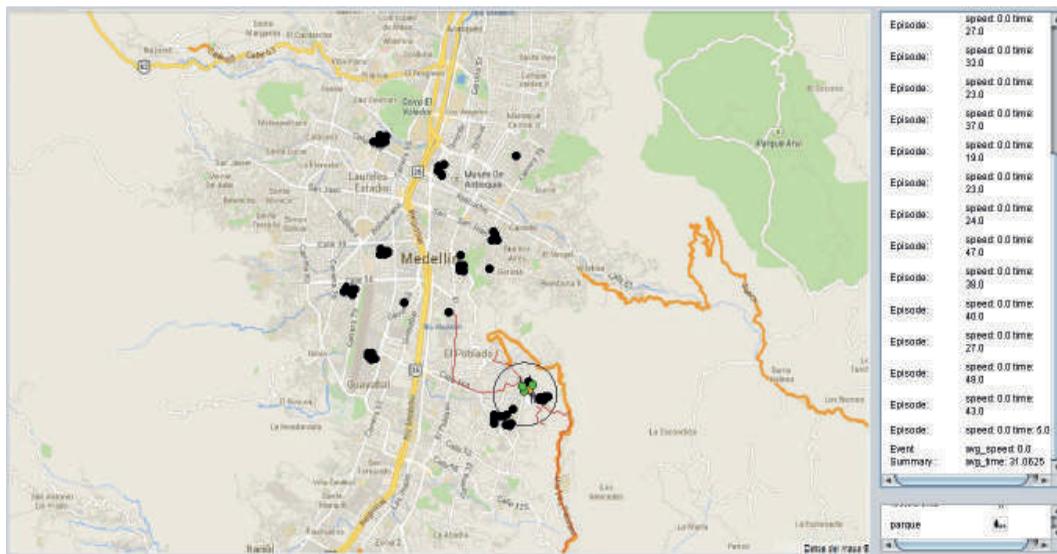


Figura 11. Resultados de la Consulta 2, marco situado en una región donde hubo episodios.

pico el lunes (cuando hay mayor congestión). Por la misma razón, el tiempo promedio de las cercanías, indicador b), fue también mayor el lunes (p. ej. debido a una congestión, un taxi permanece durante más tiempo cerca de un determinado sitio).

El valor del indicador c) para ambos días es razonable porque i) los sitios de interés

seleccionados están dispersos por la ciudad, aunque algunos están ubicados en el mismo barrio y ii) como el número de carreras analizado fue solo de 40 el domingo y 33 el lunes, esto disminuye la probabilidad de que, p. ej., dos pasajeros hayan tenido una estadía en el mismo sitio o en sitios cercanos (a menos de 100 m). El mismo razonamiento aplica para el indicador d). El indicador e) sugiere, p.

ej., que un pasajero que se dirigió a un hotel pasó durante su carrera cerca de dos hoteles. Como todos los hoteles seleccionados eran de categoría similar (hoteles de 3 y 4 estrellas), esto podría ayudar a detectar hoteles similares al hotel inicialmente considerado. Además, si se dispone de datos sobre congestión de las vías en tiempo real (aplicaciones como Waze, <https://www.waze.com>), un cambio de hotel podría ser conveniente para el pasajero ya que se podría evitar una congestión y la carrera sería más corta (ahorro para el pasajero). Análogamente, el indicador f) podría ayudar a detectar restaurantes alternativos al inicialmente considerado.

En un segundo experimento, se consideraron 120 trayectorias (de la mañana de un mismo día) de vehículos (buses) que circulan a través de los barrios de Rio de Janeiro (Brasil). El número total de observaciones fue 321 600, i. e., un promedio de 2680 observaciones por trayectoria. El tiempo promedio entre dos observaciones consecutivas fue dos segundos. Así, la duración promedio de cada trayectoria fue de hora y media. Los datos están disponibles en <http://www.inf.ufsc.br/~vania/dados/test-dataset.zip>.

Los sitios interesantes fueron aquellos donde las trayectorias se mantenían con velocidad cercana a cero durante más de 3 minutos (posiblemente paraderos o sitios

de congestión). Este proceso se hizo por separado al prototipo. Con base en estos sitios se obtuvieron los episodios de estadía donde Min. Stop Time = 3 min y para los episodios de tipo cercanía Min. Closeness Time = 30 seg y Min. Closeness Dist = 200 m. Se obtuvieron 5933 episodios, 1818 de estadía y 4115 de cercanía. Esto es razonable porque durante una trayectoria un bus usualmente hace varias paradas (en promedio 15 en paraderos previamente establecidos). Por otro lado, el número de cercanías fue alto (más del doble con relación a las estadías) porque cada vez que el vehículo hace una estadía se generan dos cercanías (una al acercarse y otra al alejarse de ella). Además, cuando un bus hace una estadía que está cerca (a menos de 200 m de otro paradero) se genera también una cercanía.

La Tabla 3 muestra algunos de los resultados consolidados y en la Figura 12 se presenta una muestra visual de los resultados (el círculo tenue en el centro que cubre varios puntos indica el marco de la lupa). El indicador d) sugiere que alrededor (a menos de 300 m) del recorrido de un bus ocurren en promedio 32 estadías (posiblemente de otros buses), i. e, cuando un bus hace una parada en un sitio s, es muy probable que al menos otros dos buses tengan una estadía cercana (a menos de 300 m) a s durante el transcurso de la mañana. Esto podría ayudar a analizar el grado de congestión y a la ubicación de los paraderos con el fin de disminuir la congestión.

Tabla 3. Resultados consolidados del experimento 2.

Indicador	
a) Tiempo promedio por trayectoria	90 min
b) Tiempo promedio de las cercanías	37 seg
c) Número promedio de estadías por trayectoria	15.15
d) Número promedio de estadías por trayectoria que ocurrieron cerca (a menos de 300 m) de otra estadía	32.3

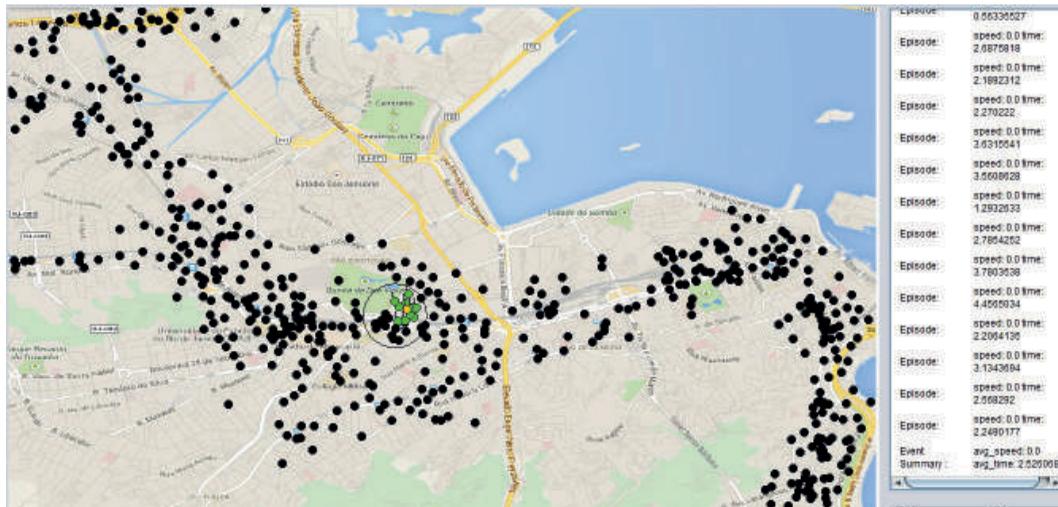


Figura 12. Muestra de los resultados del experimento 2.

4. TRABAJOS RELACIONADOS

Para facilitar la identificación de los trabajos relacionados, estos se separaron en dos categorías: trabajos orientados a la visualización de trayectorias y trabajos orientados a la visualización de BD.

4.1. VISUALIZACIÓN DE TRAYECTORIAS

Baltzer et al. [10] proponen operadores OLAP para agrupar trayectorias (Group by overlap, Group by intersection y Group by overlap and intersection) y presentan una herramienta para la visualización de sus resultados. Leonardi et al. [20] definen una BDT y un marco de trabajo [21] donde las trayectorias son tratadas como medidas y abordan el problema de su agregación y visualización, p. ej. "¿cuántas trayectorias pasaron por una determinada zona?". Por otro lado, la visualización simultánea de múltiples trayectorias puede dar lugar a despliegues confusos y atestados. Por ello, Demšar et al. [22] proponen un método

para facilitar su visualización y la identificación de patrones espacio-temporales.

Bakshev et al. [23] proponen un marco de trabajo para consultas de agregación de trayectorias (p. ej., "mostrar todas las trayectorias que pasaron por el centro de una ciudad") y su correspondiente visualización en 3D. Biadgilgn, Blok y Huisman [24] proponen cuatro métodos para visualizar una trayectoria: en un solo mapa, en varios mapas que muestran gradualmente la evolución temporal de la trayectoria, animación y en 3D mediante un cubo espacio-temporal. En Sack et al. [25] se presenta un compendio de trabajos sobre representación, análisis y visualización de trayectorias de objetos móviles.

Noyon [26] propone una representación de trayectorias para facilitar la identificación de interacciones entre dos entidades móviles al analizar sus distancias y velocidades relativas. En Parent et al. [27] se discute sobre visualización semántica de trayectorias, entre otros aspectos.

4.2. VISUALIZACIÓN DE BD

ADVIZOR [28] ofrece varias técnicas de visualización para explorar bases de datos multidimensionales. El sistema ofrece además un mecanismo visual para ejecutar drill down, el cual se asemeja a la lupa aquí propuesta pero está orientado a datos alfanuméricos.

Map cube es un operador OLAP propuesto por Shekhar et al. [29] como una extensión espacial del operador data cube [30]. Dadas n columnas de agrupamiento, map cube genera en un mapa los subtotales de las medidas numéricas y espaciales (estas son agregadas mediante la función unión geométrica) para todas las posibles combinaciones de estas columnas, i. e., 2^n subtotales. Moreno et al. [31] extendieron este operador con diferentes funciones de agregación espacial y en [32] Moreno et al. lo extendieron para la generación de trayectorias a partir de un conjunto de hechos (p. ej. el orden cronológico de los crímenes cometidos por un mismo individuo genera una trayectoria). Sin embargo, el operador no identifica episodios ni considera eventos ni sitios.

Polaris [33] es una herramienta de exploración visual interactiva que facilita el análisis de BD. El usuario puede subir y bajar a través de las jerarquías de las dimensiones (operadores OLAP roll up y drill down). Por ejemplo, considérese la consulta "mostrar en rojo y azul los estados de USA que tuvieron una mayoría de votos republicanos y demócratas en el 2000, respectivamente". El analista puede hacer zoom y observar los resultados por condado en cada estado. Un sistema similar es SOVAT [34], en el que por medio de un lenguaje gráfico se pueden formular consultas como la anterior.

En ambos sistemas, esta capacidad de analizar los datos a un nivel más detallado (según la jerarquía espacial) es análoga a la de la lupa, pero ésta última se enfoca en el análisis de trayectorias (y de las relaciones de proximidad entre sus episodios y sitios), mientras que estos lo hacen en el análisis de datos numéricos.

Interactive OLAP [35] usa una técnica de visualización en la que cada barra horizontal representa un nivel de una dimensión. Cada barra se divide en celdas mediante líneas verticales. Cada celda corresponde a un valor del nivel (instancia) y al número de hechos asociados con este. El sistema soporta drill down, p. ej., desplegar la barra de trimestres de un año. Este aumento en el nivel de detalle se asemeja a la acción de la lupa pero está orientado a datos numéricos y no a trayectorias.

5. CONCLUSIONES Y TRABAJO FUTURO

La caída en los precios y la consecuente masificación de dispositivos móviles con equipos de rastreo ofrecen la oportunidad de recopilar grandes volúmenes de datos sobre objetos móviles. La visualización de estos datos es un aspecto clave para facilitar su análisis y para ayudar a los analistas en la detección de posibles patrones espacio-temporales. Para tal fin, en este artículo se propuso una herramienta de visualización de trayectorias. La herramienta propuesta, una lupa espacio-temporal, facilita la visualización de trayectorias en una región determinada. Esto permite detectar, p. ej., la relación de cercanía de una trayectoria con otras o con los sitios a su alrededor. De hecho, en el artículo se proponen y definen formalmente las relaciones de cercanía derivadas entre

trayectorias y entre trayectorias y sitios. Una contribución adicional del artículo fue el enriquecimiento del modelo de Campora et al. [17] con elementos que permiten la formulación de consultas más expresivas y que contribuyen también en el aspecto de visualización de la herramienta.

En cuanto a los experimentos, aunque se trabajó con un volumen de datos pequeño y se requieren pruebas más exhaustivas, estos evidenciaron algunos patrones espacio-temporales incipientes, p. ej., se pudo observar que en promedio una carrera de un taxi, cuyo destino final es un hotel, pasa cerca de dos hoteles similares (de la misma categoría en estrellas). Si esta información estuviese disponible para el pasajero, p. ej., en su celular (en una aplicación similar a Waze) o a través del GPS que muchos taxis incluyen, él se podría decidir por uno de estos hoteles y así acortar la carrera con el consiguiente ahorro de dinero.

En cuanto a trabajos futuros, se consideran los siguientes: i) Permitir que el analista defina el tamaño del marco de la lupa. En este momento el tamaño del marco es fijo, lo que impide enfocarse en regiones más grandes. ii) Detectar otros tipos de relaciones entre las trayectorias y los eventos y sitios. Por ejemplo, se podría analizar si existen trayectorias de objetos móviles que recurrentemente visitan, i. e., tienen episodios de estadía en ciertos tipos de sitios, p. ej., objetos que suelen incluir siempre en su recorrido sitios de entretenimiento. También se pueden incluir otros umbrales en el nivel Event/Site para ayudar en el análisis de trayectorias, p.ej. un umbral máximo de congestión. iii) Incluir visualizaciones de proyecciones probabilísticas (p. ej. predecir dónde estará un objeto dado que estuvo en cierto sitio). Un punto de partida para este aspecto es [36-37]. Y iv) convertir el prototipo en una aplicación web e integrarlo con tecnologías como

Google Maps y KML (Keyhole Markup Language, <https://developers.google.com/kml>).

REFERENCIAS BIBLIOGRÁFICAS

- [1] Kimball, R., Ross, M., Thornthwaite, W., Mundy, J. & Becker, B. (2008). The data warehouse lifecycle toolkit. Nueva York: Wiley, p.672.
- [2] Saylor, M. (2013). The mobile wave: how mobile intelligence will change everything. Nueva York, EE.UU.: Vanguard Press, p.304.
- [3] Bolstad, P. (2012). GIS fundamentals: A first text on geographic information systems. San Francisco, EE.UU.: Eider Press, p.674.
- [4] Spaccapietra, S., Parent, C., Damiani, M. L., Fernandes de Macêdo, J.A., Porto, F. & Vangenot, C. (2008). A conceptual view on trajectories. *Data & Knowledge Engineering*, 65(1), pp.126-146.
- [5] Hägerstrand, T. (1970). What about people in regional science? *Papers of the Regional Science Association*, 24(1), pp.6-21.
- [6] Orlando, S., Orsini, R., Raffaetà, A. & Roncato, A. (2007). Trajectory data warehouses. *Journal of Computing Science and Engineering*, 1(2). pp.211-232.
- [7] Deng, K., Xie, K., Zheng, K. & Zhou, X. (2011). Trajectory indexing and retrieval. En *Computing with spatial trajectories* (pp.35-60). Boston, MA, EE.UU.: Springer.

- [8] Bogorny, V., Heuser, C.A. & Alvares, L.O. (2010). A conceptual data model for trajectory data mining. *Lecture Notes in Computer Science*, 6292, pp.1-15.
- [9] Braz, F., Orlando, S., Orsini, R., Raffaetà, A., Roncato, A. & Silvestri, C. (2006). Towards a trajectory data warehouse (CS-2006-4). Mestre, Italia: Università di Venezia Ca' Foscari, Dip. Informatica.
- [10] Baltzer, O., Dehne, F., Hambrusch, S. & Rau-Chaplin, A. (2008). OLAP for trajectories. 19th International Conference on Database and Expert Systems Application. Turín, Italia.
- [11] Orlando, S., Orsini, R., Raffaetà, A., Roncato, A. & Silvestri, C. (2007). Spatio-temporal aggregations in trajectory data warehouses. En *Data warehousing and knowledge discovery* (pp. 66-77). Berlin, Alemania: Springer.
- [12] Cuzzocrea, A. & Mansmann, S. (2009). OLAP visualization: models, issues, and techniques. En *Encyclopedia of data warehousing and mining* (pp. 1439-1446). Hershey, EE.UU.: IGI Publishing.
- [13] Jeung, H., Lung Yiu, M. & Jensen, C.S. (2011). Trajectory pattern mining. En *Computing with spatial trajectories* (pp. 143-177). Berlín, Alemania: Springer.
- [14] New York City Department of Transportation. (2012). *New York City screenline traffic flow 2011*. Nueva York, EE.UU.
- [15] Department for Transport. (2012). *Annual road traffic estimates 2011*. Londres, Inglaterra.
- [16] Jiang, B., Li, Z. (2005). Geovisualization: design, enhanced visual tools and applications. *The Cartographic Journal*, 42(1), pp.3-4.
- [17] Campora, S., Fernández de Macedo, J.A. & Spinsanti, L. (2011). St-Toolkit: a framework for trajectory data warehousing. 14th AGILE Conference on Geographic Information Science. Utrecht, Holanda.
- [18] Malinowski, E., Zimányi, E. (2008). Advanced data warehouse design: from conventional to spatial and temporal applications. Berlín, Alemania.: Springer, p.444.
- [19] Carver, R.P. (1976). The case against statistical significance testing. *Harvard Educational Review*, 48(3), pp.378-399.
- [20] Leonardi, L., Marketos, G., Frentzos, E., Gitrakos, N., Orlando, S., Pelekis, N., ... Theodoridis, Y. (2010). T-Warehouse: visual OLAP analysis on trajectory data. *IEEE 26th International Conference on Data Engineering*. Long Beach, CA, EE.UU.
- [21] Leonardi, L., Orlando, S., Raffaetà, A., Roncato, A., Silvestri, C., Andrienko, G. & Andrienko, N. (2013). A general framework for trajectory data warehousing and visual OLAP. *Geoinformatica*, 18(2), pp.273-312.
- [22] Demšar, U. & Virrantaus, K. (2010). Space-time density of trajectories: exploring spatio-temporal patterns in movement data. *International Journal of Geographical Information*, 10(10), pp.1527-1542.
- [23] Bakshev, S., Spisanti, L., Fernández de Macedo, J.A., Vidal, V. & Casanova, M. (2011). Semantic visualization

- of trajectories. 13th International Conference on Enterprise Information Systems. Beijing, China.
- [24] Biadgilgn, D., Blok, C. & Huisman, O. (2011). Assessing the cartographic visualization of moving objects. *Momona Ethiopian Journal of Science*, 3(22), pp.5-12.
- [25] Sack, J., Speckmann, B., Van Loo, E. & Weibel, R. (2011). Representation, analysis and visualization of moving objects. 2nd Seminar on Representation, Analysis and Visualization of Moving Objects. Dagstuhl, Alemania.
- [26] Noyon, V., Claramunt, C. & Devogele, T. (2007). A relative representation of trajectories in geographical spaces. *GeoInformatica*, 11(4), pp.479-496.
- [27] Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., ... Yan, Z. (2013). Semantic trajectories modeling and analysis. *ACM Computing Surveys*, 45(4), pp. 39-75.
- [28] Eick, S.G. (2000). Visualizing multi-dimensional data. *ACM SIGGRAPH Computer Graphics*, 34(1), pp.61-67.
- [29] Shekhar, S., Lu, C.T., Tan, X. & Chawla, S. (2001). Map cube: a visualization tool for spatial data warehouses. En *Geographic data mining and knowledge discovery* (pp.73-108). Nueva York, EE.UU.: Taylor and Francis.
- [30] Agarwal, S., Agrawal, R., Deshpande, P.M., Gupta, A., Naughton, J.F., Ramakrishnan, R. & Sarawagi, S. (1996). On the computation of multidimensional aggregates. *International Conference on Very Large Databases*. Mumbai, India.
- [31] Moreno, F., Arango, F. & Fileto, R. (2009). Extending the map cube operator with multiple spatial aggregate functions and map overlay. 17th International Conference on Geoinformatics. Fairfax, Virginia, EE.UU.
- [32] Moreno, F., Alzate, J.C. & Hernández, J. (2014). Extracting trajectories with the map cube operator. *Applied Mathematics & Information Sciences*, 8(5), pp.2149-2160.
- [33] Stolte, C., Tang, D. & Harahan, P. (2002). Query, analysis, and visualization of hierarchically structured data using Polaris. 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Edmonton, Canadá.
- [34] Scotch, M. & Parmanto, B. (2005). SOVAT: Spatial OLAP visualization and analysis tool. 38th Hawaii International Conference on System Sciences, Hawaii, EE.UU.
- [35] Techapichetvanich, K., Datta, A. (2005). Interactive visualization for OLAP. *Lectures Notes on Computer Science*, 3482, pp.206-214.
- [36] Liu, C., Jou, E. & Lee, C. (2010). Analysis and prediction of trajectories using bayesian network. 6th International Conference on Natural Computation. Yantai, Shandong, China.
- [37] Dore, A. & Regazzoni, C. (2010). Interaction analysis with a bayesian trajectory model. *Intelligent Systems*, 25(3), pp.32-40