

Coincidencia de mapas para trayectorias GPS de baja frecuencia de muestreo

Yin Lou

Microsoft Research Asia
v-yilou@microsoft.com

Chengyang Zhang

Microsoft Research Asia
v-chenz@microsoft.com

Yu Zheng

Microsoft Research Asia
yuzheng@microsoft.com

Xing Xie

Microsoft Research Asia
xingx@microsoft.com

Wei Wang

Universidad de Fudan
weiwang1@fudan.edu.cn

Yan Huang

Universidad del norte de Texas
huangyan@unt.edu

RESUMEN

Coincidencia de mapas es el proceso de alinear una secuencia de posiciones de usuario observadas con la red de carreteras en un mapa digital. Es un paso previo al procesamiento fundamental para muchas aplicaciones, como la gestión de objetos en movimiento, el análisis del flujo de tráfico y las direcciones de conducción. En la práctica, existe una gran cantidad de trayectorias GPS de baja tasa de muestreo (por ejemplo, un punto cada 2-5 minutos). Desafortunadamente, la mayoría de los enfoques actuales de correspondencia de mapas solo se ocupan de datos GPS de alta frecuencia de muestreo (generalmente un punto cada 10-30 s) y se vuelven menos efectivos para puntos de baja frecuencia de muestreo a medida que aumenta la incertidumbre en los datos. En este artículo, proponemos un novedoso algoritmo global de correspondencia de mapas llamado *ST-Matching* para trayectorias GPS de baja tasa de muestreo. *ST-Matching* considera (1) las estructuras espaciales geométricas y topológicas de la red de carreteras y (2) las limitaciones temporales / de velocidad de las trayectorias. Basado en el análisis espacio-temporal, un *gráfico candidato* se construye a partir de la cual se identifica la mejor secuencia de ruta coincidente. Comparamos el *ST-Matching* con el algoritmo incremental y el algoritmo global de emparejamiento de mapas basado en la distancia media de Fréchet (AFD). Los experimentos se realizan tanto en conjuntos de datos sintéticos como reales. Los resultados muestran que nuestro algoritmo de coincidencia *ST* supera significativamente al algoritmo incremental en términos de precisión de coincidencia para trayectorias de muestreo bajo. Mientras tanto, en comparación con el algoritmo global basado en AFD, *ST-Matching* también mejora la precisión y el tiempo de ejecución.

Descriptores de Sujeto y Categorías

H.2.8 [Aplicaciones de base de datos]: Bases de datos espaciales y SIG.

Términos generales

Algoritmos, Diseño

Palabras clave

Coincidencia de mapas, GPS, trayectoria, red de carreteras

1. INTRODUCCIÓN

En los últimos años se ha visto un aumento espectacular de los sistemas de guía de viaje portátiles o montados en el tablero y los sistemas de GPS integrados

El permiso para hacer copias digitales o impresas de todo o parte de este trabajo para uso personal o en el aula se otorga sin cargo siempre que las copias no se hagan o distribuyan con fines de lucro o ventaja comercial y que las copias lleven este aviso y la cita completa en la primera página. . Copiar de otra manera, o volver a publicar, publicar en servidores o redistribuir a listas, requiere un permiso específico previo y / o una tarifa.

ACM GIS '09, 4-6 de noviembre de 2009. Seattle, WA, EE. UU. (C) 2009 ACM ISBN 978-1-60558-649-6 / 09/11 ... \$ 10,00

PDA y teléfonos inteligentes. La proliferación de estos dispositivos ha permitido la recopilación de una gran cantidad de trayectorias GPS. Cada vez más aplicaciones, como planificador de rutas [7], buscador de rutas en caliente [16], análisis de flujo de tráfico [15], red social geográfica [23], han comenzado a utilizar información de datos GPS para lograr una mejor calidad de los servicios.

Normalmente, una trayectoria de GPS consta de una secuencia de puntos con información de latitud, longitud y marca de tiempo. Sin embargo, estos datos no son precisos debido a errores de medición causados por la limitación de los dispositivos GPS y el error de muestreo causado por la frecuencia de muestreo. [17]. Por lo tanto, las posiciones GPS observadas a menudo deben alinearse con la red de carreteras en un mapa digital dado. Este proceso se llama *correspondencia de mapas*. La correspondencia de mapas es un paso previo al procesamiento fundamental para muchas aplicaciones basadas en trayectorias, como la gestión de objetos en movimiento, el análisis del flujo de tráfico y las direcciones de conducción. La dificultad de la correspondencia de mapas puede diferir enormemente según la precisión del GPS y la frecuencia de muestreo.

Este documento aborda el problema de *error de muestreo*. En particular, en la práctica existe una gran cantidad de *baja frecuencia de muestreo* (por ejemplo, un punto cada 2 minutos) Trayectorias GPS. Son datos registrados por la aplicación recopilados a partir de consultas ad-hoc basadas en la ubicación o generados en los escenarios donde se desea ahorrar en costos de energía y costos de comunicación. Por ejemplo, hay Más de 60.000 taxis en Beijing, entre los cuales muchos están integrados con GPS. Dado que los taxistas viajan con mucha frecuencia, la frecuencia de muestreo debe reducirse para ahorrar consumo de energía y lograr un tiempo de respuesta razonable. Desafortunadamente, los enfoques actuales de emparejamiento de mapas solo tratan con datos de GPS de alta frecuencia de muestreo (generalmente un punto cada 10-30 s) y se vuelven menos efectivos para puntos de baja frecuencia de muestreo a medida que aumenta la incertidumbre en los datos.

La mayoría de los enfoques de correspondencia de mapas existentes emplean *local* o *incremental* algoritmos que mapean *Actual* o *vecino* posiciones en segmentos de carreteras vectoriales en un mapa. Para un enfoque que solo considera las posiciones actuales, el resultado se ve muy afectado por los errores de medición. La precisión es generalmente baja porque la correlación de los puntos vecinos se pasa por alto por completo. El algoritmo de coincidencia incremental en [8] [7] persigue la coincidencia local de una pequeña parte de la trayectoria. Al hacer coincidir una nueva posición, se consideran su posición anterior y el último borde coincidente. Aunque rápido en el cálculo, el rendimiento de este enfoque es sensible a la disminución de la frecuencia de muestreo.

Por otro lado, un *global* el algoritmo se alinea *trayectoria completa* con la red de carreteras. En términos generales, un enfoque global logra una mayor precisión a un costo computacional más alto. Existente global

Los algoritmos de emparejamiento se basan todos en ciertas medidas de distancia / similitud (como la distancia de Fréchet promedio propuesta en [3]). La falta de verdad del terreno (es decir, la trayectoria "verdadera" del objeto en movimiento) dificulta la evaluación de la calidad de coincidencia real. Además, los algoritmos de coincidencia global actuales solo emplean análisis espacial, mientras que descuidan las limitaciones temporales / de velocidad de las trayectorias. Esto también los hace vulnerables a la disminución de la frecuencia de muestreo.

En este artículo nos referimos a *baja frecuencia de muestreo* como un punto cada 2 minutos o más. Con tal frecuencia de muestreo, la distancia entre dos puntos puede alcanzar más de 1300 m, incluso la velocidad de un vehículo es de solo 40 km / h! Esto plantea un gran desafío para el problema de la correspondencia de mapas porque a medida que aumenta la distancia entre dos puntos vecinos, se puede usar menos información para deducir las ubicaciones precisas del objeto. El problema se agrava cuando un objeto en movimiento viaja a gran velocidad o hay muchas intersecciones entre dos puntos vecinos observados.

Para abordar el desafío, proporcionamos dos observaciones clave que conducen a nuestro enfoque, como se ilustra en los siguientes ejemplos.

Observación 1: Los verdaderos caminos tienden a ser directos, en lugar de rotundos.

Ejemplo 1: Considere la trayectoria GPS de un taxi visualizada en la Figura 1. El taxi viaja de norte a sur. , y están tres puntos de muestreo consecutivos.



Figura 1. Ilustración de la observación 1

La mayoría de los algoritmos de coincidencia de mapas coincidirían con el GPS encerrado en un círculo observación ya su tramo de carretera más cercano, es decir, el vertical. Sin embargo, desde su posición anterior y posición futura, nosotros puede determinar que debe coincidir con el segmento de carretera horizontal porque es poco probable que este taxi tome una rotonda viaje primero a la carretera horizontal y luego regrese a la vertical. Esto implica que la información topológica de la red de carreteras se puede combinar con el contexto de la posición para proporcionar mejores resultados de coincidencia.

Observación 2: los caminos verdaderos tienden a seguir las limitaciones de velocidad del camino.

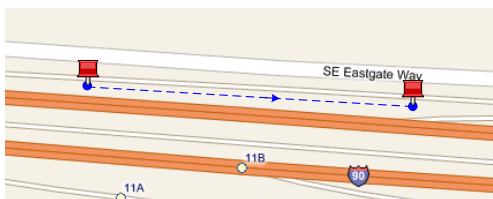


Figura 2. Ilustración de la observación 2

Ejemplo 2: Considere otra trayectoria GPS de taxi visualizada en la Figura 2. Sin información de velocidad, es casi imposible saber si estos dos puntos pertenecen a la carretera o al servicio.

camino. Sin embargo, si calculamos la velocidad promedio de esta ruta como 80km / h Basándonos en la distancia de los dos puntos y sus marcas de tiempo, podemos determinar que es muy probable que los dos puntos estén en la carretera. Esto implica que la información temporal / de velocidad también es útil en el proceso de emparejamiento.

Sobre la base de las observaciones anteriores, proponemos un novedoso algoritmo de comparación de mapas global llamado *ST-Matching* para trayectorias GPS de baja frecuencia de muestreo en este documento. Hasta donde sabemos, nuestro trabajo es el primer algoritmo de correspondencia de mapas que incorpora tanto (1) las estructuras espaciales geométricas y topológicas de la red de carreteras y (2) las limitaciones temporales / de velocidad de las trayectorias. Combinando el análisis espacial y temporal, se construye un gráfico candidato a partir del cual el algoritmo de coincidencia ST intenta encontrar una ruta de coincidencia global con la puntuación más alta.

En resumen, este trabajo realiza las siguientes aportaciones:

Proponemos un novedoso algoritmo de coincidencia global llamado STmatching para trayectorias GPS de baja frecuencia de muestreo. Combina características espaciales y temporales que no se exploran con métodos anteriores.

Realizamos extensos experimentos en conjuntos de datos tanto sintéticos como reales. El conjunto de datos real se recopila del mundo físico y está etiquetado por personas reales. Por lo tanto, puede evaluar mejor el rendimiento de los algoritmos de correspondencia de mapas que los datos sintéticos utilizados en la mayoría de los trabajos existentes.

Nuestro algoritmo se evalúa en términos de tiempo de ejecución y precisión de coincidencia. Los resultados muestran que nuestro algoritmo de coincidencia ST supera significativamente al algoritmo incremental en términos de precisión de coincidencia para trayectorias de baja frecuencia de muestreo. Mientras tanto, en comparación con el algoritmo global basado en AverageFréchet-Distance, ST-Matching también mejora la precisión y el tiempo de ejecución.

El resto del artículo está organizado de la siguiente manera. La sección 2 analiza el trabajo relacionado. Los preliminares y la definición del problema se presentan en la Sección 3. La descripción general del sistema se da en la Sección 4. Luego, el algoritmo de emparejamiento ST se propone en la Sección 5 con una discusión y un análisis detallados tanto en la Sección 5 como en la 6. La Sección 7 presenta los resultados de la evaluación experimental. Finalmente, concluimos el artículo con la Sección 8.

2. TRABAJO RELACIONADO

En esta sección, presentamos el trabajo relacionado con el problema de correspondencia de mapas. Dado que nuestro algoritmo se basa en el cálculo de la ruta más corta, también revisaremos brevemente el problema de la ruta más corta.

2.1 Problema de correspondencia de mapas

Hay varios estudios sobre la correspondencia de las observaciones de GPS en un mapa digital. Estos enfoques se pueden clasificar generalmente en tres clases: método local / incremental [4] [8] [20], método global [1] [3] [21] y método estadístico [11] [18].

Los métodos locales / incrementales intentan encontrar *local* coincidencia de geometrías. El método incremental en [8] utiliza dos medidas de similitud para evaluar los bordes candidatos, una para la similitud de distancia y la otra para la similitud de orientación. La medida de similitud combinada se calcula como la suma de las puntuaciones individuales. La complejidad del tiempo es una vez que encontramos los bordes adyacentes para cada muestra, donde es el número de puntos GPS que se van a emparejar. El método de "recorte adaptativo" en [20] utiliza el algoritmo de Dijkstra para construir la ruta más corta en el gráfico de espacio libre local. Corre en

\log tiempo, donde myn son el número de bordes en la red de carreteras y el número de puntos GPS, respectivamente. Además del emparejamiento incremental, [4] propone un método de emparejamiento basado en segmentos para asignar valores de confianza para diferentes puntos de muestreo. Primero, hace coincidir los segmentos de alta confianza y, luego, los segmentos de baja confianza utilizando bordes previamente emparejados. En general, cuando coincide con una nueva posición, un método local / incremental solo considera una pequeña parte de la trayectoria que está cerca de la posición. Funciona rápido y funciona bien cuando la frecuencia de muestreo es muy alta (por ejemplo, 2-5 segundos). Sin embargo, a medida que disminuye la frecuencia de muestreo, el problema de "salto de arco" [8] se vuelve prominente, lo que provoca una degradación significativa de la precisión. Por el contrario, con un aumento razonable de la complejidad del tiempo, nuestro algoritmo de coincidencia ST global es más robusto a la disminución de la frecuencia de muestreo.

Los métodos globales apuntan a igualar el *trayectoria completa* con la red de carreteras. Paper [21] emplea un método de ajuste fuera de línea que tiene como objetivo encontrar una ruta de peso mínimo en función de la distancia de edición. Otros métodos [1] [3] se basan en la distancia de Fréchet o sus variantes. La distancia de Fréchet tiene en cuenta la continuidad de las curvas y, por lo tanto, es adecuada para comparar trayectorias. En [1], el algoritmo aplica una búsqueda paramétrica sobre todos los valores críticos. Luego resuelve el problema de decisión al encontrar un camino monótono en el espacio libre desde la esquina inferior izquierda hasta la esquina superior derecha. Corre en (registro_2) tiempo, donde *metro* y *norte* son el número de bordes y el número de nodos de la red de carreteras, respectivamente. Este trabajo se amplía en [3] con la distancia media de Fréchet para reducir el efecto de los valores atípicos. El papel [3] también utiliza una distancia de Fréchet débil que se ejecuta en (\log) tiempo con una calidad de coincidencia similar. Los métodos globales tienen como objetivo minimizar la distancia de Fréchet entre la trayectoria y los segmentos de carretera coincidentes. Sin embargo, los métodos actuales no han intentado utilizar rutas verdaderas para evaluar la precisión de coincidencia real. Mientras tanto, la información temporal / de velocidad en las trayectorias generalmente se pasa por alto. Como demuestran nuestros experimentos, cuando se aplica a datos reales de baja frecuencia de muestreo con rutas verdaderas etiquetadas por humanos, la coincidencia ST supera a los métodos basados en la distancia de Fréchet tanto en términos de precisión de coincidencia como en tiempo de ejecución.

Los modelos estadísticos también se utilizan para hacer coincidir las observaciones de GPS. En [18], se presenta un método de correspondencia de mapas basado en un clasificador bayesiano que incorpora un modelo de Markov oculto para modelar las restricciones topológicas de la red de carreteras. En [11] se propone un método mejorado basado en el filtro de Kalman extendido y la interpolación spline cúbica. Los enfoques estadísticos parecen particularmente efectivos para manejar los errores de medición del GPS, como lo demuestran sus experimentos. Pueden combinarse con nuestro método de compatibilidad con ST para lograr mejores resultados.

2.2 Cálculo de la ruta más corta

Ahora revisamos brevemente el trabajo relacionado sobre el cálculo de la ruta más corta, ya que se utiliza en nuestro algoritmo ST-Matching. Esta es un área que ha recibido una gran atención de investigación durante décadas. Se puede encontrar una buena encuesta en [5].

El algoritmo básico en el cálculo de la ruta más corta es el algoritmo de Dijkstra. En la práctica, el algoritmo A* [10] se utiliza a menudo como una alternativa más eficiente. Un algoritmo * utiliza una función heurística para guiar la búsqueda hacia el destino. Otras estrategias como la búsqueda bidireccional [12], la descomposición de la búsqueda [13] y la búsqueda jerárquica [14] también se utilizan a menudo en aplicaciones reales.

Se pueden agregar algunos pasos de preprocesamiento para acelerar la búsqueda de la ruta más corta. Por ejemplo, los algoritmos ALT en [6] emplean el

combinación de puntos de referencia y desigualdad triangular para alcanzar un límite inferior más estrecho que la distancia euclidiana utilizada en el algoritmo A*. La poda basada en el alcance [9] es otro método para calcular el límite inferior con fines de poda.

3. DECLARACIÓN DEL PROBLEMA

En esta sección, daremos los preliminares y definiremos formalmente el problema de la correspondencia de mapas para trayectorias GPS de baja frecuencia de muestreo.

Definición 1 (Registro GPS): Un registro de GPS es una colección de GPS puntos $= \{1, 2, \dots\}$. Cada punto GPS \in contiene latitud $.$, longitud $.$ y marca de tiempo $.$, como se ilustra en la parte izquierda de la Figura 3.

Definición 2 (Trayectoria GPS): Una trayectoria de GPS es una secuencia de puntos GPS con el intervalo de tiempo entre cualquier Puntos GPS que no superan un cierto umbral $.$, es decir $: 1 \rightarrow 2 \rightarrow \dots \rightarrow$, donde \in , y $0 < +1. - < \Delta (1 \leq <)$. La Figura 3 muestra un ejemplo de trayectoria GPS. es el intervalo de muestreo. En este artículo, nos enfocamos en la baja tasa de muestreo. Trayectorias GPS con $\Delta \geq 2$.

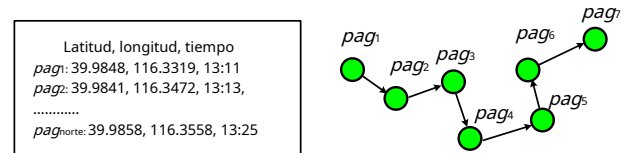


Figura 3. Ilustración del registro GPS y la trayectoria GPS

Definición 3 (Segmento de carretera): Un tramo de carretera es un borde dirigido que está asociado con una identificación $.$, una velocidad de viaje típica $.$, un valor de longitud $.$, un punto de partida $.$, un punto final $.$ y una lista de puntos intermedios que describe la carretera mediante una polilínea. La Figura 4 muestra varios segmentos de carreteras reales en Bing Map Search [2]. Tenga en cuenta que una carretera puede contener varios tramos de carretera.



Figura 4. Ilustración del segmento de la carretera

Definición 4 (Red de carreteras): Una red de carreteras es un gráfico dirigido $(,)$, donde es un conjunto de vértices que representan las intersecciones y los puntos terminales de los tramos de carretera, y es un conjunto de bordes que representan tramos de carretera.

Definición 5 (Camino): Dados dos vértices $.$ en una red de carreteras $.$, un sendero es un conjunto de tramos de carretera conectados que comienzan en $.$ y termina en $.$, es decir $: 1 \rightarrow 2 \rightarrow \dots \rightarrow$, donde $1. = , . = +1. , 1 \leq < .$

Ahora, el problema de la correspondencia de mapas se define como:

Dada una trayectoria de GPS sin procesar y una red de carreteras $(,)$, encuentra el camino de eso combina con su camino real.

4. DESCRIPCIÓN GENERAL DEL SISTEMA

La arquitectura de nuestro sistema de correspondencia de mapas propuesto se muestra en la Figura 5. Se compone de tres componentes principales: *Preparación de candidatos*, *análisis espacial y temporal*, y *Resultado Pareo*.

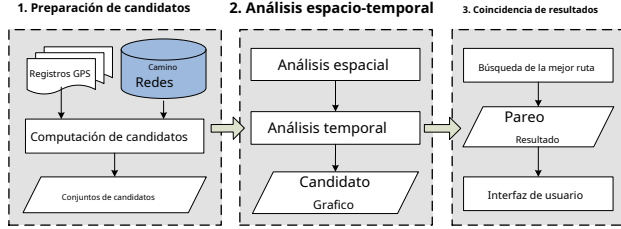


Figura 5. Descripción general de la arquitectura del sistema

Preparación de candidatos Este componente contiene una base de datos de la red de carreteras con información indexada sobre bordes y vértices. Acepta una trayectoria GPS sin procesar dada del usuario y luego recupera todos los posibles puntos candidatos para cada punto de muestreo en la trayectoria. Este paso se puede realizar de manera eficiente con el índice espacial basado en cuadrícula incorporado. El resultado de este componente es un conjunto de puntos candidatos y los tramos de carretera candidatos en los que se encuentran.

Análisis espacial y temporal Este componente realiza un análisis espacial seguido de un análisis temporal de los conjuntos de candidatos recuperados y la trayectoria que se va a emparejar.

El análisis espacial no solo considera la distancia entre un único punto GPS y los tramos de carretera candidatos para este punto, sino que también tiene en cuenta la información topológica de la red de carreteras. Para evitar caminos indirectos, empleamos el camino más corto para medir la similitud entre cada camino candidato y el camino "verdadero".

El análisis temporal mide la velocidad de viaje promedio real entre los puntos vecinos. Luego compara la velocidad promedio con las restricciones de velocidad típicas en cada ruta candidata. La información se puede utilizar posteriormente para hacer coincidir la trayectoria con la ruta candidata con la mayoría de las condiciones de velocidad similares durante ese intervalo de tiempo.

Después del análisis espacial y temporal, se construye un gráfico candidato como resultado de este componente. Los nodos del gráfico son el conjunto de puntos candidatos para cada observación GPS, y los bordes del gráfico son el conjunto de las rutas más cortas entre dos puntos candidatos vecinos cualesquiera. A los nodos y bordes se les asignan valores de peso basados en los resultados del análisis espacial / temporal.

Coincidencia de resultados Este componente evalúa el gráfico candidato utilizando la información de peso asignada durante el análisis espacial / temporal. Hace coincidir la trayectoria dada con la ruta con la puntuación más alta en el gráfico de candidatos. Luego, los resultados se visualizan en una interfaz que se puede adaptar a diferentes dispositivos de usuario final. Los resultados también se pueden almacenar en una base de datos de tráfico para admitir aplicaciones externas como la gestión del tráfico o las direcciones de conducción.

5. El algoritmo de emparejamiento ST

En esta sección, describimos nuestro algoritmo ST-Matching en detalle.

5.1 Preparación de candidatos

Trayectoria dada $= p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_n$, primero recuperamos un conjunto de tramos de carretera candidatos dentro del radio r de cada punto p_i , $1 \leq i \leq n$. Luego calculamos los puntos candidatos, que son segmentos de línea proyecciones de p_i a estos tramos de carretera definidos de la siguiente manera.

Definición 6 (Proyección de segmento de línea): La proyección del segmento de línea de un punto p_i a un tramo de carretera e es el punto q en e semejante que $q = \arg \min_{q \in e} \|p_i - q\|$, donde $\| \cdot \|$ devuelve la distancia entre p_i y cualquier punto q en e .

En el resto de este documento, usamos p_i y q respectivamente para denotar la i -ésima ventaja del candidato y el punto candidato de p_i . Como se muestra en la Figura 6, El candidato p_i puede ser p_1, p_2, p_3 .

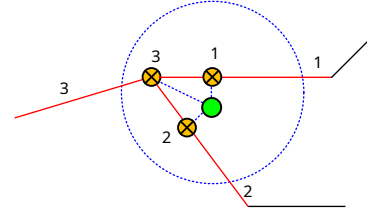


Figura 6 Puntos candidatos para un punto de muestreo

Para facilitar la búsqueda de puntos candidatos, la red de carreteras se indexa mediante una cuadrícula. Una vez que se recuperan los conjuntos de puntos candidatos para todos los puntos de muestreo en la trayectoria, el problema es cómo elegir un candidato de cada conjunto para que los mejores partidos $p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_n$ sean los mejores.

5.2 Análisis espacial

En el análisis espacial, utilizamos información tanto geométrica como topológica de la red de carreteras para evaluar los puntos candidatos encontrados en el paso anterior. La información geométrica se incorpora mediante *probabilidad de observación*, y la información topológica se expresa utilizando *probabilidad de transmisión*.

Definición 7 (Probabilidad de observación): La probabilidad de observación se define como la probabilidad de que un muestreo GPS en un punto p_i coincide con un punto candidato q calculado sobre la base de la distancia entre los dos puntos p_i y q .

En términos generales, el error en una medición GPS puede describirse razonablemente como una distribución normal. (p_i, q) de la distancia entre p_i y q . Indica la probabilidad de que un GPS observación p_i puede coincidir con un punto candidato q en la realidad carretera sin considerar sus puntos vecinos. Formalmente, nosotros definir la probabilidad de observación $O(p_i, q)$ de p_i wrt q como:

$$O(p_i, q) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(p_i - q)^2}{2\sigma^2}\right) \quad (1)$$

dónde $\sigma = 20$ es la distancia entre p_i y q . En este artículo, usamos una distribución normal de media cero con un estándar desviación de 20 metros basada en evaluación empírica.

La probabilidad de observación no tiene en cuenta el contexto de la posición de un punto GPS. Esto a veces conduce a resultados de coincidencia incorrectos. La figura 7 muestra un ejemplo de este tipo. Las líneas gruesas representan una carretera y la línea vertical delgada representa una carretera local.

Aunque p_i está más cerca de q en la carretera local, deberíamos

partido a 2 en la carretera si ya sabemos que es vecinos -1 y +1 están en la autopista. Esto se basa en la observación de que es poco probable que un vehículo tome una ruta rotonda. (*Observación 1*).

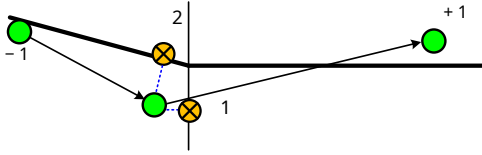


Figura 7. Un ejemplo que necesita probabilidad de transmisión

Para caracterizar la intuición anterior, calculamos el camino más corto entre dos puntos candidatos vecinos -1 y +1. Entonces nosotros definir *probabilidad de transmisión* como sigue:

Definición 8 (Probabilidad de transmisión): Dado dos candidatos puntos -1 y +1 para dos puntos de muestreo GPS vecinos -1 y +1 respectivamente, la probabilidad de transmisión de -1 a +1 es definida como la probabilidad de que el camino "verdadero" desde -1 a +1 siga el camino más corto desde -1 a +1.

Calculamos la probabilidad de transmisión como

$$-1 \rightarrow +1 = \frac{-1 \rightarrow +1}{-1, \rightarrow (+1)} \quad (2)$$

dónde $-1 \rightarrow +1$ es la distancia euclidiana entre y_{-1}, y_{+1} , $-1, \rightarrow (+1)$ es la longitud del camino más corto desde -1 a +1.

Combinando la ecuación (1) y (2), definimos el **análisis espacial función** $-1 \rightarrow +1$ como la *probabilidad de observación* y *probabilidad de transmisión*:

$$-1 \rightarrow +1 = \text{función de probabilidad de observación} \times \text{probabilidad de transmisión}, 2 \leq \dots \leq \dots \quad (3)$$

dónde -1 y +1 son dos puntos candidatos para dos puntos GPS vecinos -1 y +1 respectivamente.

La ecuación (3) calcula la probabilidad de que un objeto se mueva desde -1 a +1 utilizando el producto de dos funciones de probabilidad, por lo tanto, la información geométrica y topológica se toman en consideración. Tenga en cuenta que, en la práctica, es poco probable que un objeto en movimiento siga siempre la ruta estrictamente más corta. Por lo tanto, los *probabilidad de observación* no se puede omitir de (3).

Con análisis espacial, para dos puntos GPS vecinos -1 y +1, un conjunto de caminos candidatos $-1 \rightarrow +1$ son generados. Cada A la ruta se le asigna un valor de medición espacial calculado a partir de la Ecuación (3).

5.3 Análisis temporal

El análisis espacial puede distinguir la ruta real de las otras rutas candidatas en la mayoría de los casos. Sin embargo, hay situaciones que el análisis espacial por sí solo no podría manejar.

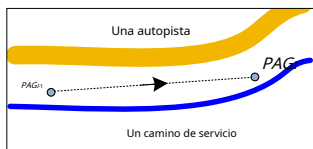


Figura 8. Un ejemplo que necesita un análisis temporal

Considere el ejemplo que se muestra en la Figura 8. La línea gruesa es una carretera y las líneas delgadas son carreteras de servicio cercanas a la carretera. La función de análisis espacial puede producir el mismo valor si dos puntos -1 y +1 coinciden con la autopista o la vía de servicio. Sin embargo, si calculamos la velocidad media a partir de -1 a +1 como 80 km/h , los relacionaríamos con la carretera considerando los límites de velocidad de la vía de servicio (*Observación 2*).

Más formalmente, dados dos puntos candidatos -1 y +1 para dos puntos de muestreo GPS vecinos -1 y +1 respectivamente, el camino más corto desde -1 a +1 se denota como una lista de carreteras segmentos $[s_1, s_2, \dots, s_n]$. La *velocidad media* $-1, \rightarrow (+1)$ de El La ruta más corta se calcula de la siguiente manera:

$$-1, \rightarrow (+1) = \frac{1}{\Delta -1 \rightarrow +1} \quad (4)$$

dónde Δ es la longitud de $\Delta -1 \rightarrow +1$ es el intervalo de tiempo entre dos puntos de muestreo y -1 . Tenga en cuenta que cada segmento de la carretera también está asociado con un valor de velocidad típico v_i . Empleamos la distancia del coseno para medir la similitud entre la velocidad media real de -1 a +1 y las limitaciones de velocidad del camino. Considere el vector que contiene elementos del mismo valor $-1, \rightarrow (+1)$ y el vector $(v_1, v_2, \dots, v_n)^T$. La *función de análisis temporal* se define de la siguiente manera.

$$-1 \rightarrow +1 = \frac{1 \times (-1, \rightarrow (+1))}{(v_1, v_2, \dots, v_n)^T \times (-1, \rightarrow (+1))} \quad (5)$$

Como en la función de análisis espacial, -1 y +1 son dos candidatos cualesquiera puntos por -1 y +1 respectivamente.

5.4 Coincidencia de resultados

Con el análisis espacial y temporal anterior, estamos listos para describir nuestro *ST-Matching* algoritmo utilizado en el componente de coincidencia de resultados.

En general, después del análisis espacial y temporal, podemos generar un gráfico candidato (C, E) por trayectoria: $-1 \rightarrow \dots \rightarrow +1$, es un conjunto de puntos candidatos para cada punto de muestreo GPS, y E es un conjunto de bordes que representan los caminos más cortos entre dos puntos candidatos vecinos, como se muestra en la Figura 9. Cada nodo en C está asociado con $-1 \rightarrow +1$. Cada borde está asociado con $-1 \rightarrow +1$.

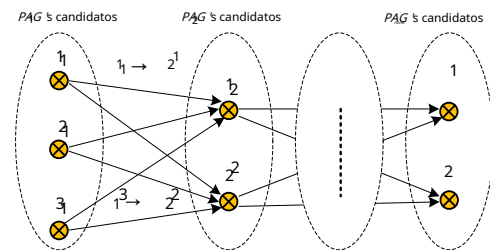


Figura 9. El gráfico candidato (C, E)

Combinando la ecuación (3) y (5), definimos **Función ST** por

$$-1 \rightarrow +1 = \text{función de probabilidad de observación} \times \text{función de análisis temporal}, 2 \leq \dots \leq \dots \quad (6)$$

Una secuencia de ruta candidata para toda la trayectoria es un camino en el gráfico candidato, denotado como $1 \rightarrow 2 \rightarrow \dots \rightarrow n$. La puntuación para tal secuencia candidata es $P = \sum_{i=1}^{n-1} p_{i,i+1}$. De todas las secuencias candidatas, nuestro objetivo es encontrar la que tenga la puntuación general más alta como la que mejor coincida camino para la trayectoria. Más formalmente, la mejor ruta de coincidencia PAG por una trayectoria se selecciona como:

$$P = \max_{i \in \{1, 2, \dots, n\}} P_i \quad (7)$$

El algoritmo 1 describe el marco del algoritmo ST-Matching. Primero calculamos los conjuntos de puntos candidatos para cada punto de muestreo GPS en T . Luego construimos el gráfico candidato G_c basado en análisis espacial y temporal. Finalmente, el algoritmo informa la secuencia de la ruta con el valor de función ST más alto de G_c como el resultado.

Algoritmo 1 Entrada de algoritmo de coincidencia ST: red de carreteras R , una trayectoria $1 \rightarrow 2 \rightarrow \dots \rightarrow n$.

Resultado: la secuencia emparejada $1 \rightarrow 2 \rightarrow \dots \rightarrow n$ en T .

```

1: inicializar  $C$  como una lista vacía; // una lista de puntos
candidatos
2: para  $i = 1$  hacer 3:
     $C_i = \text{GetCandidates}(R, p_i, r)$ ; //candidatos dentro del radio
4:  $G_c = \text{ConstructGraph}(C)$ ; //gráfico de construcciones
5:  $P = \text{FindMatchedSequence}(G_c)$ ;
6: devuelve  $P$ ;

```

La *FindMatchedSequence* procedimiento tiene como objetivo encontrar el más largo camino en G_c . En general, encontrar la ruta más larga en un gráfico es un NPC problema. Sin embargo, G_c es un gráfico acíclico dirigido (DAG) y el La respuesta se puede calcular de manera eficiente utilizando el orden topológico del gráfico. Es fácil ver que tal orden topológico puede obtenerse de la construcción de G_c . El algoritmo 2 muestra detalles del procedimiento *FindMatchedSequence*.

Algoritmo 2 FindMatchedSequence

Entrada: gráfico de candidatos G_c .

Salida: la secuencia más larga $1 \rightarrow 2 \rightarrow \dots \rightarrow n$ en G_c .

```

1: deja  $P$  denotar la puntuación más alta calculada hasta el
momento; 2: deja  $P_i$  denotar el padre del candidato actual;
3: para cada  $c \in C$  hacer
4:  $P[c] = 0$ ;
5: por  $i = 2$  a  $n$  hacer
6: para cada  $c \in C_i$  hacer
7:  $max = -\infty$ ;
8: para cada  $p \in C_{i-1}$  hacer
9:  $alt = P[p] + p_{p,c}$ ;
10: Si ( $alt > max$ ) luego
11:  $max = alt$ ;
12:  $P_i[c] = p$ ;
13:  $P[c] = max$ ;
14: inicializar  $C$  como una lista vacía; // una lista de puntos coincidentes en
orden inverso
15:  $C = C_n$ ;
dieciséis: por  $i = n-1$  a 2 hacer
17:  $C_i = C_i \cup C_{i+1}$ ;
18:  $C_i = C_i$ ;
19:  $C_i = C_i \cup \{c_i\}$ ;
20: regreso  $P$ ;

```

La Figura 10 muestra un ejemplo en ejecución del procedimiento anterior. El valor de la probabilidad de observación, la probabilidad de transmisión y la función de medición temporal también se enumeran en la figura.

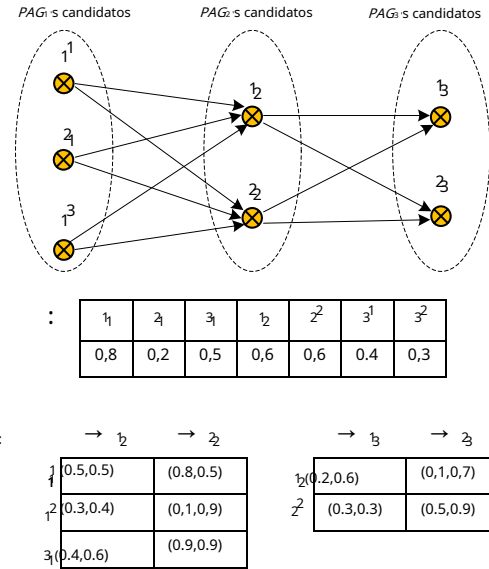


Figura 10. Un ejemplo de *FindMatchedSequence* procedimiento

El algoritmo primero llena la tabla P como sigue.

1^1	2^1	3^1	1^2	2^2	3^2	3^3
0,8	0,2	0,5				

A continuación, considere candidato 1^1 en la 1^1 fila y 2^2 en la 2^2 columna. $P_{1^1,2^2} = 0,6 \times 0,5 \times 0,5 = 0,15$. Similar, $P_{1^1,3^2} = 0,072$, $P_{2^1,3^2} = 0,144$. Por lo tanto $P_{1^1} = \max\{0,8 + 0,15, 0,2 + 0,072, 0,5 + 0,144\} = 0,95$, y $P_{1^1} = 1$. El padre es 1^1 . Repetimos el proceso anterior para todos los candidatos y completar la tabla P .

1^1	2^1	3^1	1^2	2^2	3^2	3^3
0,8	0,2	0,5	0,95	1,04	1,076	1,175

Encontramos que 3^3 tiene el puntaje general más alto, y nosotros escoger 3^3 como el resultado coincidente del último punto GPS. Basado en la información principal almacenada, el algoritmo *FindMatchedSequence* informará el emparejamiento resultado como $1 \rightarrow 2 \rightarrow 3$.

5.5 Localización de estrategias de coincidencia de ST

El algoritmo ST-Matching propuesto es un algoritmo global, ya que solo podemos identificar la mejor secuencia de ruta después de calcular la puntuación general para toda la trayectoria. En la práctica, cuando una trayectoria tiene demasiados puntos (es decir, es muy grande), o se desea un procesamiento en línea, podemos localizar el algoritmo ST-Matching mediante la construcción un gráfico candidato parcial G_c sobre una ventana corrediza del trayectoria, donde w es el tamaño de la ventana.

Como se ilustra en la Figura 11, cada gráfico candidato parcial G_c es construido a partir de un subconjunto de trayectoria T . Encontramos la mejor secuencia de ruta coincidente para este subconjunto de manera similar a como ocurre en STmatching global. Luego deslizamos la ventana y repetimos el proceso.

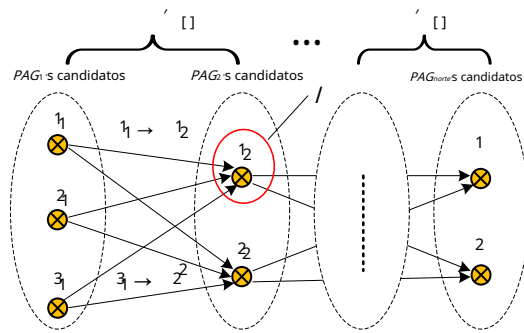


Figura 11. Localización de ST-Matching (= , =)

La partición de la trayectoria en ventanas puede reducir el retraso promedio y ahorrar espacio de almacenamiento para el procesamiento en línea, pero no necesariamente acelera el tiempo de procesamiento general porque la parte más costosa del algoritmo ST-Matching es el cálculo de la ruta más corta. Para reducir la complejidad del cálculo, podemos retener heurísticamente los puntos candidatos con top puntuación hasta ahora, reduciendo así el número de caminos más cortos que deben calcularse para el siguiente punto de muestreo. Cuando $\epsilon = 1$, ST-Matching degenera en un algoritmo incremental.

6. ANÁLISIS DE ALGORITMO

Ahora analizamos la complejidad temporal de nuestro algoritmo STMatching propuesto. Dejar n denotar el número de puntos de muestreo en la trayectoria dada, y m denotar el número de tramos de carretera en la red de carreteras. Suponemos además que el número máximo de candidatos de un punto de muestreo es k .

Observe que el número de caminos más cortos en k es $(k-1) \cdot 2$. Por lo tanto, la complejidad de tiempo para construir el gráfico candidato es $O(k^2)$. Para el *FindMatchedSequence* procedimiento, cada borde en G se visita exactamente una vez. El tiempo la complejidad es $O(nk^2)$. Por lo tanto, nuestro algoritmo global ST-Matching tiene la complejidad de tiempo de $O(k^2 \log n + k^2)$.

Tenga en cuenta que para cualquier punto de muestreo dado, prácticamente podemos elegir un valor pequeño de ϵ . Por lo tanto, la complejidad del tiempo está cerca de $\log n$. Por el contrario, el algoritmo global en [7] tiene la complejidad de $O(n^2)$ usando la distancia de Fréchet y $O(n^2)$ utilizando la Distancia de Fréchet Débil.

En la práctica, Dijkstra no suele ser una buena opción para calcular la ruta más corta. Podemos elegir el algoritmo A* [10] o ALT [6] para lograr un mejor tiempo de ejecución. Un algoritmo A* utiliza estimaciones de distancias al destino para guiar la selección de vértices en una búsqueda desde la fuente. Para lograr un tiempo de ejecución aún más rápido, el algoritmo ALT emplea la combinación de puntos de referencia y desigualdad triangular para obtener un límite inferior más estricto, pero requiere más espacio en disco para el preprocesamiento. Además, las estrategias de localización propuestas en la Sección 5.5 también son útiles para reducir el tiempo de respuesta y mejorar la eficiencia de los cálculos.

7. EXPERIMENTOS

En esta sección, primero presentamos la configuración experimental, incluido el conjunto de datos que usamos y algunos parámetros que seleccionamos en el experimento. A continuación, describimos los enfoques de evaluación que aplicamos. Luego informamos los principales resultados con algunas discusiones. Nuestros resultados experimentales se comparan tanto con el algoritmo incremental como con la distancia de Fréchet promedio (AFD)

algoritmo de coincidencia de mapas global basado. Los resultados se evalúan principalmente en términos de tiempo de ejecución y precisión de coincidencia.

7.1 Configuración experimental

7.1.1 Descripción del conjunto de datos

Red de carreteras: En nuestros experimentos, usamos la red de carreteras de Beijing como se visualiza en la Figura 12. El gráfico de red contiene 58,624 vértices y 130,714 tramos viales.



Figura 12. Red de carreteras de Beijing

Datos de trayectoria sintética: Los datos sintéticos utilizados en nuestros experimentos contienen trayectorias generadas aleatoriamente por nuestro propio simulador. El simulador funciona de la siguiente manera. Primero selecciona aleatoriamente dos vértices en la red de carreteras y calcula la parte superior caminos más cortos entre ellos. Luego selecciona aleatoriamente una trayectoria desde el K caminos como la verdad fundamental, denotado como $1, 2, \dots, K$. La motivación detrás de esto es que los objetos en movimiento generalmente siga la dirección desde el origen al destino, pero no necesariamente siga estrictamente el camino más corto. Tenga en cuenta que el intervalo de tiempo entre dos puntos vecinos no es uniforme. Para recuperar una trayectoria con el intervalo de muestreo deseado, el simulador selecciona un segmento de carretera de cada K segmentos en K , es decir, selecciona

$1, 1 + \frac{1}{K}, \dots, 1 + \frac{K-1}{K}$, donde $\frac{1}{K} = \frac{1}{K}$. El ajuste del muestreo

Por lo tanto, la tasa se logra cambiando el valor de K . El simulador genera un punto GPS con información de marca de tiempo estimada para cada segmento de carretera seleccionado. Los puntos se producen para seguir la distribución normal de media cero con la desviación estándar de 20 metros.

Datos de trayectoria real: A diferencia de los enfoques existentes, también utilizamos etiquetas humanas *verdadero camino* datos como el **verdad fundamental** en nuestros experimentos. De cientos de trayectorias reales recopiladas del sistema GeoLife [22], seleccionamos 28 trayectorias que varían en cobertura espacial, longitud, intervalo de tiempo y número de tramos de carretera. Se etiquetan manualmente con las rutas verdaderas. La Figura 13 muestra el histograma de las características reales del conjunto de datos.

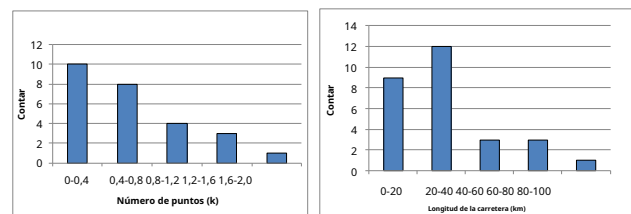


Figura 13. Histograma de un conjunto de datos reales

La Figura 14 muestra un ejemplo de trayectoria real visualizada usando Bing Map API. Las líneas azules conectadas por marcadores verdes representan la trayectoria del GPS, y las líneas rojas conectadas por marcadores rojos representan las verdaderas rutas del usuario.

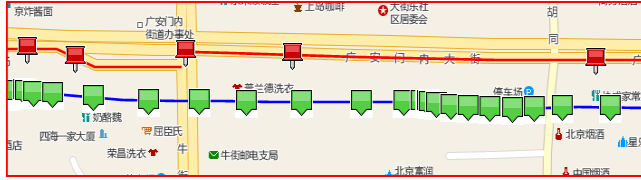


Figura 14. Un ejemplo de datos de trayectoria etiquetados en GeoLife [22]

7.1.2 Selección de parámetros

Tasa de muestreo de datos: Este artículo se centra en las trayectorias GPS de baja frecuencia de muestreo. El intervalo de muestreo varía de 2 minutos a casi 6 minutos para datos sintéticos y varía de 30 segundos a 5 minutos para datos de trayectoria real.

Parámetros para el algoritmo de coincidencia ST: En nuestros experimentos, elegimos $\delta = 5$ empíricamente como el número máximo de candidatos de cualquier punto de muestreo. El radio de búsqueda de candidatos se selecciona como $r = 100$. Para la estimación de la probabilidad de observación, utilizamos una distribución normal con $\mu = 0$ y $\sigma = 20$ metro. Al evaluar estrategias localizadas, el tamaño de la ventana varía de 1 a 40, y elegimos retener la parte superior 1, 3, 5 nodos.

Parámetros para algoritmos de línea base: Comparamos nuestros experimentos con dos algoritmos de referencia en este artículo. Para el algoritmo incremental, utilizamos la configuración empírica $\delta = 10$, $\alpha = 0,17$, $\beta = 1,4$, $\gamma = 10$, y $\eta = 4$ como se sugiere en [3]. Para el algoritmo basado en AFD, establecemos $\delta = 10$.

7.2 Enfoques de evaluación

Verdad fundamental: En los datos sintéticos, la verdad fundamental se selecciona al azar entre los principales K caminos más cortos desde el origen hasta el destino. En los datos reales recopilados de GeoLife, la verdad básica se establece utilizando caminos verdaderos etiquetados por humanos. Durante el proceso de etiquetado, si un usuario no puede identificar el tramo de carretera real en el que reside el punto de muestreo, dicho punto se etiqueta con una etiqueta especial para que luego no se cuente en la evaluación.

Criterios de evaluación: Nuestro algoritmo ST-Matching se evalúa tanto en términos de tiempo de ejecución como de calidad de coincidencia. El tiempo de ejecución se mide utilizando el tiempo de ejecución real del programa. La calidad de coincidencia se mide utilizando dos métricas de precisión *Precisión por número* (P_n) y *Precisión por longitud* (P_l) definido como sigue.

$$P_n = \frac{\# \text{ de coincidencias correctas}}{\# \text{ de candidatos}} \\ P_l = \frac{\# \text{ de coincidencias correctas}}{\# \text{ de candidatos}} \times \frac{\text{Longitud de la trayectoria real}}{\text{Longitud de la trayectoria candidata}}$$

Líneas de base: Principalmente comparamos nuestro algoritmo ST-Matching con dos líneas de base: un algoritmo incremental que realiza un emparejamiento basado en el resultado de emparejamiento anterior de un punto; y un algoritmo de coincidencia global que tiene como objetivo minimizar la distancia media de Fréchet (AFD). Para evaluar los efectos del análisis temporal, desarrollamos un algoritmo "S-Matching" eliminando el análisis temporal de ST-Matching. S-Matching también se compara con el método ST-Matching original.

7.3 Resultados experimentales

7.3.1 Resultados sobre datos sintéticos

Impacto de Max. Número de puntos candidatos: Las figuras 15 y 16 muestran el impacto del número máximo de puntos candidatos en la precisión de coincidencia y el tiempo de ejecución de ST-Matching. En las cifras, "Inf" significa que no limitamos el número de candidatos.

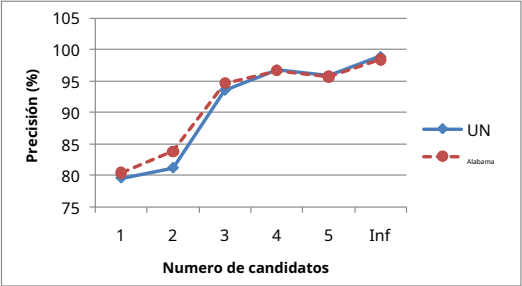


Figura 15. Precisión con respecto al número de candidatos

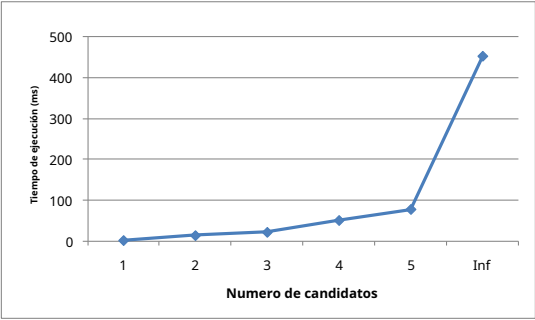


Figura 16. Tiempo de ejecución con número de candidatos

Como se muestra en la Figura 15, la precisión generalmente aumenta a medida que el algoritmo toma en consideración más puntos candidatos. Sin embargo, una gran cantidad de puntos candidatos para un punto GPS conduciría a una gran cantidad de cálculos de ruta más cortos, lo que aumentará significativamente el tiempo de ejecución. La Figura 16 muestra que si limitamos el número de candidatos a 5, el tiempo de ejecución promedio de nuestro algoritmo de correspondencia de mapas es aceptable.

Emparejamiento ST vs Emparejamiento S:

A continuación, evaluamos los efectos del análisis temporal comparando la calidad de coincidencia de ST-Matching y S-Matching wrt con el intervalo de muestreo.

En la generación de trayectorias GPS sintéticas, el cambio del intervalo de muestreo se logra mediante el ajuste del parámetro δ como se discutió en la Sección 7.1.1. Nosotros elegimos δ como 9, 11, 13, 15 y 17 respectivamente. El valor del intervalo de muestreo para cada valor de δ se enumera en la Tabla 1.

Cuadro 1. Intervalo de muestreo en datos sintéticos

δ	9	11	13	15	17		
Intervalo (min)	2,91	3,42	4,14	5,12	5,77		

Calculamos los 5 mejores caminos más cortos. Para cada δ , Generamos aleatoriamente 10 trayectorias GPS y probamos nuestros algoritmos de correspondencia de mapas. Luego tomamos la precisión promedio.

Como se muestra en la Tabla 2, la precisión generalmente disminuye a medida que disminuye la frecuencia de muestreo. Esto se debe a que la suposición de la ruta más corta ya no se mantendrá. También observamos que STMatching supera a S-Matching la mayor parte del tiempo, lo que indica que las restricciones temporales son de hecho útiles en la mayoría de los casos.

Tabla 2. Coincidencia de ST frente a Coincidencia de S (precisión)

/	Anorte		AL	
	S- Coincidencia	ST-Matching	S-Matching	ST-Matching
9	0,917	0,935	0,938	0,954
11	0,893	0,913	0,920	0,944
13	0,895	0,891	0,928	0,926
15	0,837	0,855	0,890	0,896
17	0,803	0,823	0,843	0,863

Algoritmo global basado en ST-Matching vs Incremental vs AFD

A continuación, comparamos nuestro algoritmo de coincidencia ST con el algoritmo de coincidencia de mapas global incremental y basado en AFD. Como muestran las Figuras 17 y 18, ST-Matching supera de manera significativa tanto al algoritmo incremental como al basado en ADF. Mientras tanto, notamos que el rendimiento del algoritmo incremental degenera drásticamente cuando la frecuencia de muestreo disminuye mientras que los algoritmos globales son más robustos. al cambio de frecuencia de muestreo.

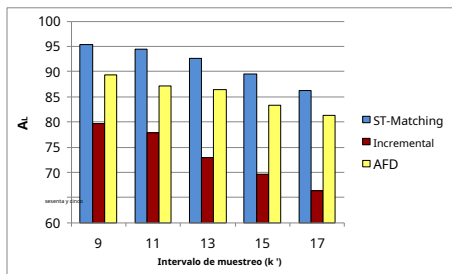


Figura 17. AL (%) intervalo de muestreo wrt

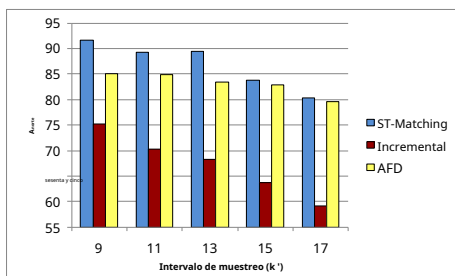


Figura 18. An (%) intervalo de muestreo wrt

7.3.2 Resultados sobre datos reales

Ahora evaluamos el algoritmo ST-Matching en datos reales recopilados de Geolife.

Algoritmo global basado en ST-Matching vs Incremental vs AFD

Las figuras 19 y 20 muestran el cambio de la calidad de coincidencia wrt a la frecuencia de muestreo en datos reales. Se puede ver claramente que nuestro algoritmo STMatching supera tanto al algoritmo incremental como al basado en AFD en datos reales. A medida que disminuye la frecuencia de muestreo, la precisión del algoritmo incremental se reduce drásticamente, mientras que dos algoritmos globales muestran nuevamente un rendimiento más consistente. También notamos que la precisión de coincidencia de dos algoritmos globales tiende a ser muy cercana cuando el intervalo de muestreo es alto. Esto implica

que las restricciones temporales se vuelven menos efectivas cuando los puntos vecinos están lejos unos de otros.

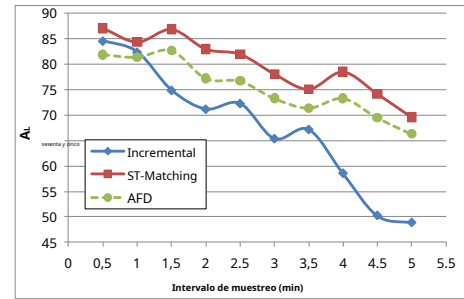


Figura 19. Un intervalo de muestreo de wrt sobre datos reales

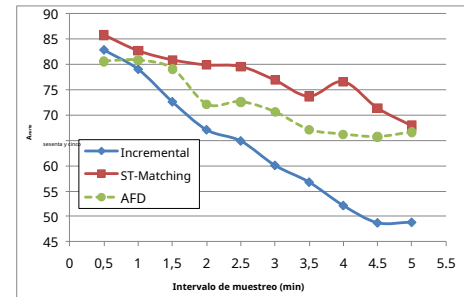


Figura 20. Anorte intervalo de muestreo wrt en datos reales

Tiempo de ejecución

Para evaluar la eficiencia del algoritmo ST-Matching, comparamos el tiempo de ejecución de nuestro algoritmo con el método global basado en AFD. Como se demuestra en la Figura 21, cuando el número de puntos es pequeño, los dos métodos exhiben un tiempo de ejecución similar. Sin embargo, a medida que aumenta el número de puntos, el tiempo de ejecución de El enfoque basado en ADF aumenta drásticamente.

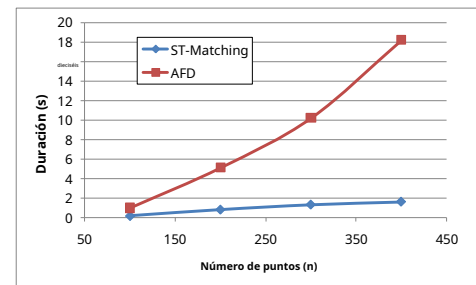


Figura 21. Comparación del tiempo de ejecución con datos reales

Estrategias de localización

Finalmente, evaluamos el efecto de las estrategias de localización propuestas en la Sección 5.5 con una trayectoria real de más de 400 puntos. En la Figura 22 se puede ver que, en general, cuando el tamaño de la ventana alcanza un cierto valor, el aumento adicional del tamaño de la ventana tiene poco efecto en la mejora de la precisión. En algunos casos, el aumento del tamaño de la ventana puede incluso tener un efecto adverso en la precisión de la coincidencia. Esto implica que el ST-Matching localizado se puede utilizar prácticamente siempre que seleccionemos un tamaño de ventana apropiado. Por otro lado, reteniendo la parte superior/Los nodos reducirán significativamente la precisión de coincidencia, especialmente cuando el tamaño de la ventana es grande.

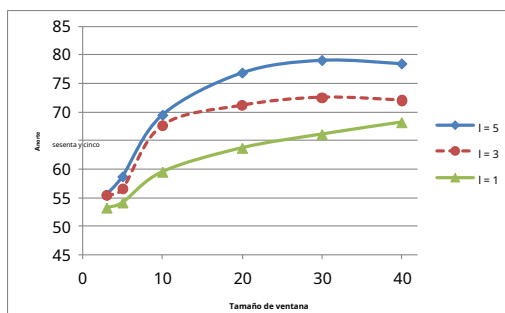


Figura 22. Evaluación de estrategias de localización

8. CONCLUSIONES

En este artículo, proponemos un nuevo algoritmo global de correspondencia de mapas llamado ST-Matching para hacer coincidir datos GPS de baja frecuencia de muestreo en un mapa digital. El algoritmo emplea análisis espacial y temporal para generar un gráfico candidato, a partir del cual se identifica una secuencia de resultados coincidentes con la suma más alta de puntuación como resultado coincidente. Los resultados del experimento demuestran que nuestro algoritmo de coincidencia ST supera significativamente al algoritmo incremental en términos de precisión de coincidencia para trayectorias de muestreo bajo. Mientras tanto, en comparación con el algoritmo global basado en ADF, ST-Matching también mejora la precisión y el tiempo de ejecución.

En nuestro trabajo futuro, planeamos construir un sistema completo de correspondencia de mapas basado en el algoritmo ST-Matching. La primera aplicación de dicho sistema será la coincidencia de trayectorias de taxi que puede ayudar en las direcciones de conducción basadas en el conocimiento del conductor de taxi.

9. REFERENCIAS

- [1] Alt, H., Efrat, A., Rote, G. y Wenk, C. 2003. Mapas planos coincidentes. En *Journal of Algorithms* 49, 2, 262-283, 2003.
- [2] Búsqueda de mapas de Bing (anteriormente búsqueda de mapas en vivo). <http://cn.bing.com/ditu/>. Corporación Microsoft.
- [3] Brakatsoulas, S., Pfoser, D., Salas, R. y Wenk, C. 2005. Sobre datos de seguimiento de vehículos con correspondencia de mapas. En *Actas de la 31a Conferencia internacional sobre bases de datos muy grandes*, 853-864, 2005.
- [4] Chawathe, Coincidencia de mapas basada en segmentos de SS. En el *Simposio de IEEE sobre vehículos inteligentes*, 2007.
- [5] Fu, L., Sun, D. y Rilett, LR Algoritmos heurísticos de ruta más corta para aplicaciones de transporte: estado del arte. En *Comput. Oper. Res.* 33, 11, 3324-3343, 2006.
- [6] Goldberg, AV y Harrelson, C. 2005. Calcular el camino más corto: Una búsqueda * se encuentra con la teoría de grafos. En *Actas del Decimosexto Simposio Anual ACM-SIAM sobre Discreto*, 156-165, 2005.
- [7] González, H., Han, J., Li, X., Myslinska, M. y Sondag, J. P. 2007. Cálculo adaptativo de la ruta más rápida en una red de carreteras: un enfoque de minería de tráfico. En *Actas de la 33a Conferencia internacional sobre bases de datos muy grandes* 794-805, 2007.
- [8] Greenfeld, J. Emparejar observaciones de GPS con ubicaciones en un mapa digital. En *Actas de la 81a Reunión Anual de la Junta de Investigación del Transporte*, 2002.
- [9] Gutman, R. Enrutamiento basado en alcance: un nuevo enfoque para los algoritmos de ruta más corta optimizados para redes de carreteras. En *Actas de 6th ALENEX*, 100-111, 2004.
- [10] Hart, N. Nilsson y B. Raphael. Una base formal para la determinación heurística de rutas de costo mínimo. En *IEEE Transactions sobre ciencia de sistemas y cibernética*, 4: 100-107, 1968.
- [11] Hummel, B. y Tischler, K. Coincidencia de mapas solo con GPS: explotación del historial de posición del vehículo, información sobre restricciones de conducción y topología de la red de carreteras en un marco estadístico. En *Proceedings of the GIS Research UK Conference*, 68-77, 2005.
- [12] JAT, Nicholson. Encontrar la ruta más corta entre dos puntos en una red. En *Computer Journal*, 9, 275-80, 1996.
- [13] JF, Dillenburg, PC, Nelson. Mejorar la eficiencia de la búsqueda mediante posibles subobjetivos. En *Modelado matemático y por computadora*, 22 (2-7): 379-414, 1995.
- [14] Jing, N., Huang, Y. y Rundensteiner, EA Optimización jerárquica de la búsqueda de rutas óptimas para aplicaciones de transporte. En *Actas de la Quinta Conferencia Internacional sobre Gestión de la Información y el Conocimiento*, 261-268, 1996.
- [15] Kuehne, R. et al. Nuevos enfoques para la gestión del tráfico en áreas metropolitanas. En el *X Simposio de IFAC sobre Control en Sistemas de Transporte*, 2003.
- [16] Li, X., Han, J., Lee, J. y González, H. Descubrimiento de rutas calientes en redes de carreteras basado en la densidad del tráfico. En el *X Simposio Internacional sobre Bases de Datos Espaciales y Temporales*, 2007.
- [17] Pfoser, D. y Jensen, CS Capturando la incertidumbre de las representaciones de objetos en movimiento. En *Actas del 6º Simposio internacional sobre avances en bases de datos espaciales*, 111-132, 1999.
- [18] Pink, O. Hummel, B. Un enfoque estadístico para la correspondencia de mapas utilizando la geometría de la red de carreteras, la topología y las restricciones de movimiento vehicular. En la *11a Conferencia Internacional IEEE sobre Sistemas de Transporte Inteligentes*, 2008.
- [19] Samet, H., Sankaranarayanan, J. y Alborzi, H. Navegación a distancia de red escalable en bases de datos espaciales. En *Actas de la Conferencia internacional ACM SIGMOD de 2008 sobre gestión de datos*, 43-54, 2008.
- [20] Wenk, C., Salas, R. y Pfoser, D. 2006. Abordar la necesidad de velocidad de coincidencia de mapas: localización de algoritmos de coincidencia de curvas globales. En *actas de la 18a Conferencia internacional sobre bases de datos científicas y estadísticas Gestión*, 2006.
- [21] Yin, H. y Wolfson, O. Un método de correspondencia de mapas basado en ocho en bases de datos de objetos en movimiento. En *Actas de la 16.a Conferencia internacional sobre gestión de bases de datos científicas y estadísticas*, 437, 2004.
- [22] Zheng, Y., Chen, Y., Xie, X. y Ma, W. GeoLife2.0: un servicio de redes sociales basado en la ubicación. En las *actas de la Conferencia Internacional sobre Gestión de Datos Móviles*, 2009.
- [23] Zheng, Y., Wang, L., Xie, X. y Ma, W. GeoLife administrando y comprendiendo su vida pasada sobre mapas. En la *9a Conferencia Internacional sobre Gestión de Datos Móviles*, 2008.