

– quantil –

# Trajectory Data Mining

Mónica Ribero

24 de noviembre de 2016

# Contenido

## 1 Introducción

## 2 Tareas

- Preprocesamiento
- Minería de Patrones en Trayectorias
- Clasificación de Trayectorias
- Transformaciones

## 3 Aplicaciones

- Aplicación a patrones de movilidad
- Aplicación a trayectorias de crimen



# Introducción

- Cantidad masiva de datos provenientes de GPS y dispositivos móviles
- objetos, personas, vehículos, animales...
- Necesidad de procesar, manejar y minar estas fuentes

# Definición

## Definition

Una **trayectoria** es la traza que deja un objeto en movimiento

$$p_1, p_2, \dots, p_n$$

donde

$$p_i = (x, y, t)$$



# Tareas

- Preprocesamiento
- Tareas de Minería
- Transformaciones



# Noise Filtering

- El ruido de los sensores y la recepción de señal afectan la calidad y precisión de los datos.
- Mean, Median Filters* toman para cada punto el promedio en ventanas de  $n$  puntos. No funciona bien cuando los puntos están muy distanciados ni con errores consecutivos

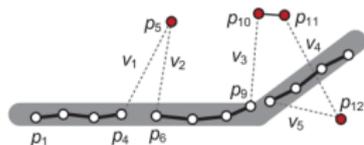


Fig. 2. Noise points in a trajectory.



# Staypoint Detection

- Detectar puntos de mayor importancia o comportamiento diferente
- El punto 1 es raro (simplemente por el ruido en la transmisión)
- Para detectar el punto 2, se establecen márgenes de desplazamiento por unidad de tiempo.

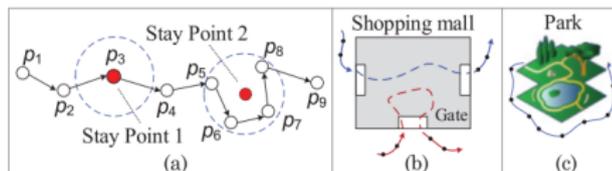


Fig. 3. Stay points in a trajectory.

# Trajectory Compression

- Reducir dimensionalidad de los datos.

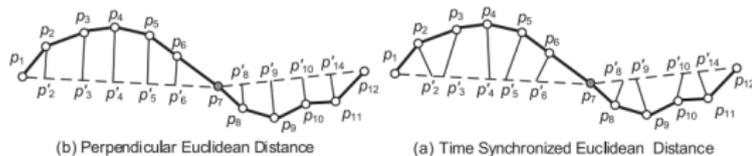


Fig. 4. Distance metric measuring the compression error.

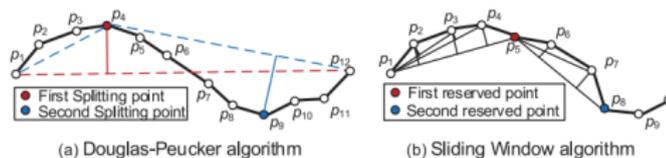


Fig. 5. Illustration of Douglas-Peucker algorithm.

# Trajectory Segmentation

- Además de disminuir la complejidad, algunas tareas requieren segmentar las trayectorias.

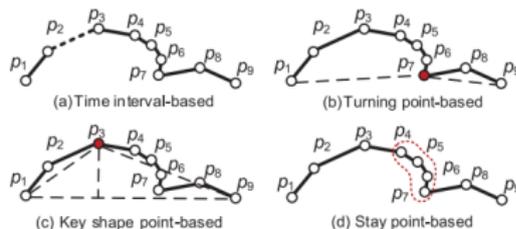


Fig. 6. Methods of trajectory segmentation.

# Map Matching

- Adaptar trayectorias a un mapa

## Subsection 2

# Minería de Patrones en Trayectorias

# Moving Together Patterns

- útiles para detectar procesos de migración, reunión, eventos de tráfico
- *Flock* (rebaño, bandada): Grupo moviéndose en un disco  $D$  por un tiempo determinado
- Convoy: Density based clustering

# Clustering

- Crear un vector de características para cada trayectoria.
- Problema: Diferentes longitudes, formas, puntos, etc.

# Clustering

- Particionar cada trayectoria en segmentos lineales.
- Construir grupos de segmentos cercanos
- Tomar un camino representativo de cada cluster

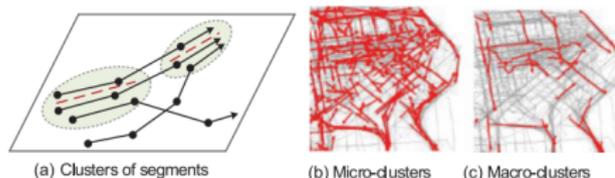
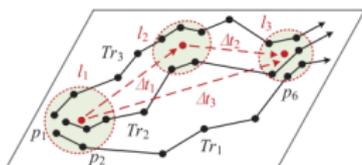


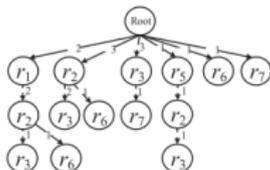
Fig. 16. Trajectory clustering based on partial segments [Li et al. 2010b].

# Patrones Secuenciales

- Detectar secuencias similares de lugares en intervalos similares de tiempo
- Métodos basados en simplificaciones por líneas
- Clustering-Based Methods: clasificar los puntos de las trayectorias en clusters de interés.



(a) Sequential trajectory patterns



(b) Suffix Tree-based sequential pattern mining

Fig. 17. Sequential pattern mining in trajectory data.

## Subsection 3

# Clasificación de Trayectorias

# Clasificación – Metodología

- Segmentar trayectorias
- Extraer características de cada segmento
- Modelo de clasificación
- Modelos de inferencia de secuencias: *Dynamic Bayesian Network* (DBN), HMM, y *Conditional Random Field* (CRF)

# Clasificación – Aplicaciones

- Aplicaciones: detectar dispositivo en movimiento, tipo de movimiento, “Ocupado, vacío, parqueado” .
- Detectar lugares de parqueo utilizando *Stay point methods*
- Utilizar estos puntos para segmentar las trayectorias
- Caracterizar cada segmento (puntos de interés, información geográfica, etc)
- Entrenar clasificador probabilístico para cada segmento
- Detectar patrones locales usando HMM

## Subsection 4

# Transformaciones







# Técnicas con Grafos

- Hyperlink-Induced Topic Search (HITS; o *hubs and authorities*)
- Algoritmo que nace en el contexto de la web para identificar buenos recomendadores (hubs) y buenas páginas (autoridades)
- En el contexto de trayectorias: hubs son usuarios y las ubicaciones las autoridades



# Trayectorias a Matrices

- Filas: usuarios
- Columnas: ubicaciones
- Entradas: número de visitas de un usuario a una ubicación
- Técnicas de *Collaborative Filtering* para completar información y hacer recomendaciones.





# Patrones de Movilidad

- “Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore”
- Shan Jiang, Joseph Ferreira, Jr., Marta C. González
- Describir el *pipeline* para captar, filtrar, analizar y sintetizar datos de celulares para interpretar la movilidad humana.
- Usar y entender esta información para la planeación.



# Metodología

- Detectar ubicaciones donde los usuarios se quedan por largos periodos.
- Detectar la casa de los usuarios
- Filtrar usuarios y seleccionar muestras representativas del CDR
- Identificar la red diaria de movilidad de los usuarios.









# Resultados

Cociente de Ubicación del motif  $m$  para comparar la distribución de  $m$  en el área con la distribución de un área de referencia (área metropolitana de Singapur)

$$LQ_m = \frac{e_m/e}{E_m/E}$$

donde

$e_m$  = Cantidad de personas en región con motif  $m$

$e$  = Cantidad de personas en región

$E_m$  = Cantidad de personas en la ciudad con motif  $m$

$E$  = Cantidad de personas en la ciudad















**GRACIAS**