

Geospatial Data Management Project

Giuseppe Maurizio Facchi
mat. 989910

Gregorio Ghidoli
mat.

A.A. 2021-2022

Indice

1	Introduzione	3
1.1	Analisi del dataset	4
1.2	Obiettivi	5
1.3	Paradigma di data mining su traiettorie	6
2	Trajectory Preprocessing	7
2.1	Stop points detection	7
2.1.1	Algoritmo basato su time-space threshold	8
2.1.2	Algoritmo basato su clustering	8
2.1.3	Algoritmo basato su modelli probabilistici	8
2.1.4	Confronto tra gli algoritmi	8
3	Trajectory Pattern Mining	9
3.1	Obiettivo 2	9
3.2	Obiettivo 3	10

1 Introduzione

Definition 1.1: Traiettorie

Una traiettoria è una traccia generata da un oggetto in movimento in spazi geografici, di solito rappresentata da una serie di punti cronologicamente ordinati

$$p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_n$$

dove ogni punto consiste in una coppia di coordinate geografiche associate ad un timestamp

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

Le traiettorie offrono la possibilità di estrarre informazioni da oggetti in movimento, come animali, veicoli o persone.

In base alla natura dei dati è possibile classificare le tipologie di dati relative a traiettorie in 4 possibili gruppi (Yu Zheng):

- **Mobilità di persone**
- Mobilità di veicoli
- Mobilità di animali
- Mobilità di fenomeni naturali

Il dataset in analisi tratta dati relativi a mobilità di persone, dove 3 visitatori di un museo hanno visto registrare **attivamente** i loro movimenti per un certo periodo di tempo con un RTLS.

Definition 1.2: RTLS

Si definisce RTLS qualsiasi sistema (hardware + software) in grado di campionare accuratamente la posizione di un'entità in uno spazio definito ad un'alta frequenza.

RTLS non è uno specifico tipo di sistema, ma più un obiettivo che può essere raggiunto utilizzando vari sistemi (Damiani)

I real-time location systems utilizzati forniscono la posizione di un tag. In base alle caratteristiche del sistema una posizione può indicare:

- **Presenza** in un'area, espressa simbolicamente
- Posizione precisa, espressa in coordinate
- Vicinanza ad un altro tag, espressa come distanza o simbolicamente

Nel caso in analisi la posizione è espressa in coordinate con CRS ESPG 3003.

Si parla in particolare di registrazione **attiva** (Yu Zheng) quando l'entità tracciata partecipa direttamente all'attività di tracciamento tramite l'utilizzo di RTLS, come nel caso delle persone in analisi.

1.1 Analisi del dataset

Il dataset fornito contiene i dati relativi alle esposizioni di un museo (Figura 1) e relativi al tracciamento di 3 persone al suo interno (Figura 2).

(Figura) (caption I dati sono stati forniti come shapefile ESRI e successivamente importati in QGIS, dando il risultato in figura)

Il tracciamento è stato effettuato per una durata di circa 10 minuti ad una frequenza di campionamento di circa 12 campioni/secondo (1 ogni 0.08 secondi). Una prima analisi è stata effettuata per ricavare le principali statistiche riguardanti le traiettorie di ogni persona, ottenendo i seguenti risultati:

Persona 57

	Distanza	Durata	Velocità	Turning Angle
Media	0.040	0.085	0.485	185.9
Mediana	0.030	0.072	0.374	182.0
Minimo	0.000156	0.0709	0.00218	0.01
Massimo	0.575	0.179	4.178	359.9

Persona 67

Media	0.052	0.089	0.613	176.8
Mediana	0.039	0.072	0.454	181.6
Minimo	0.0006	0.071	0.003	0.002
Massimo	0.493	0.179	6.226	359.9

Persona 68

Media	0.045	0.087	0.546	183.3
Mediana	0.034	0.072	0.417	183.2
Minimo	0.00078	0.071	0.010	0.006
Massimo	0.791	5.356	10.988	359.9

Tabella 1: Dati aggregati relativi alle traiettorie delle persone

Questi dati verranno utilizzati per inizializzare nel modo migliore gli algoritmi proposti.

1.2 Obiettivi

Dati le traiettorie, gli obiettivi dell'analisi sono:

- Trovare i punti di fermata delle persone (**stop points**)
- Data un'esibizione, calcolare il tempo di visita totale nei suoi pressi, considerando solo i punti di fermata
- Trovare il numero di persone che in un certo periodo temporale si trovano nei pressi della stessa esibizione

Definition 1.3: Punto di Fermata - Stop Point

Viene generalmente definito "punto di fermata", un punto nel dominio spazio-temporale dove un'entità risulta essere stazionaria. Esistono tipicamente due condizioni di stazionarietà:

- L'entità rimane ferma nello stesso punto per un certo periodo di tempo
- Definita un'area, l'entità rimane all'interno di essa per un certo periodo di tempo

1.3 Paradigma di data mining su traiettorie

L'operazione relativa all'estrarre informazioni che arricchiscano la conoscenza del dominio di un problema prende il nome di *data mining*. Una possibile organizzazione di tutte le operazioni presenti in letteratura è specificata nella seguente figura, che presenta un paradigma per il *trajectory data mining*.

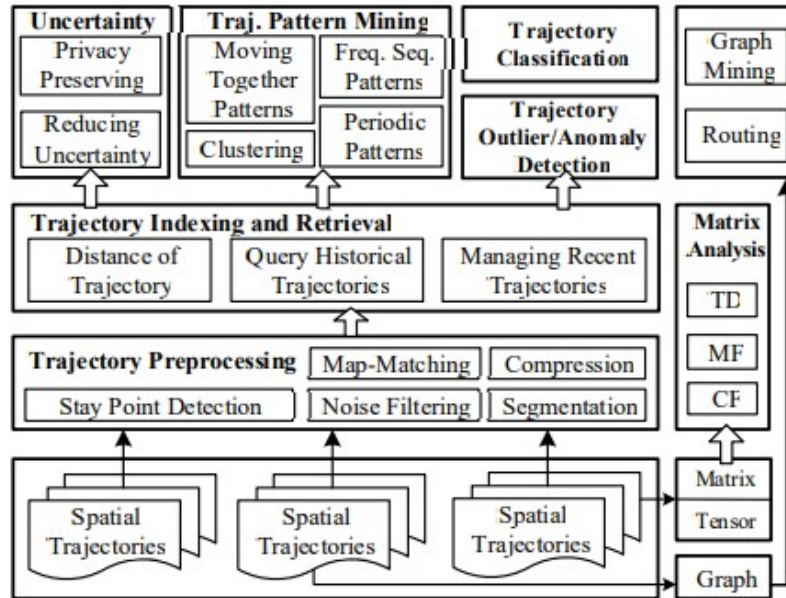


Figure 1 Paradigm of trajectory data mining

2 Trajectory Preprocessing

L'operazione di stop point detection e l'operazione di segmentazione della traiettoria fanno parte di quelle operazioni eseguite prima di effettuare ulteriori analisi sul dataset che vengono chiamate di *trajectory preprocessing*:

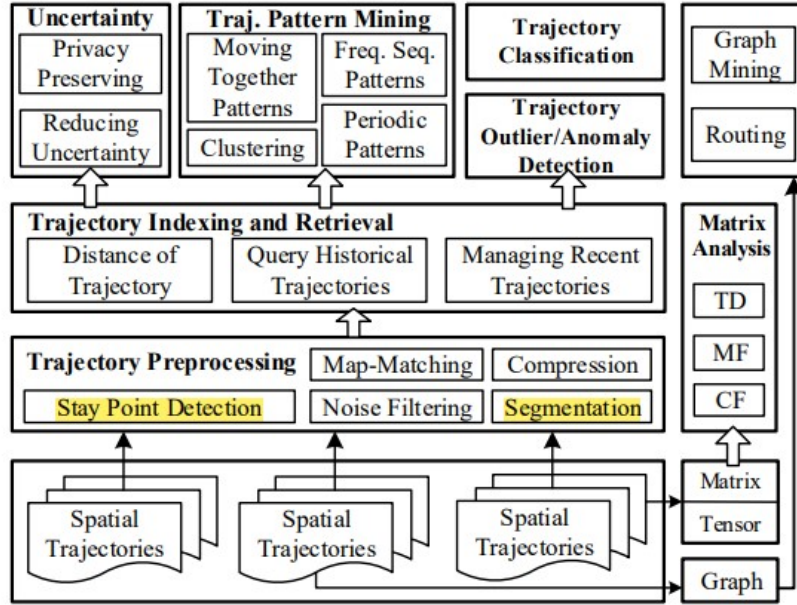


Figure 1 Paradigm of trajectory data mining

2.1 Stop points detection

In letteratura sono presenti diversi algoritmi per l'individuazione dei punti di fermata e conseguentemente per la segmentazione della traiettoria in diversi gruppi di stop. Per questo dataset sono stati presi in considerazione 3 algoritmi che seguono 3 diversi approcci:

- Algoritmo basato su time-space threshold
- Algoritmo basato su clustering
- Algoritmo basato su modelli probabilistici

- 2.1.1 Algoritmo basato su time-space threshold
- 2.1.2 Algoritmo basato su clustering
- 2.1.3 Algoritmo basato su modelli probabilistici
- 2.1.4 Confronto tra gli algoritmi

3 Trajectory Pattern Mining

Un importante punto di analisi riguarda il trajectory pattern mining, ovvero l'estrazione di pattern che arricchiscano l'informazione contenuta in una traiettoria. Lo scoprire gruppi di persone che si muovono insieme per un certo periodo di tempo ne è un tipico esempio e prende il nome di *moving together pattern*:

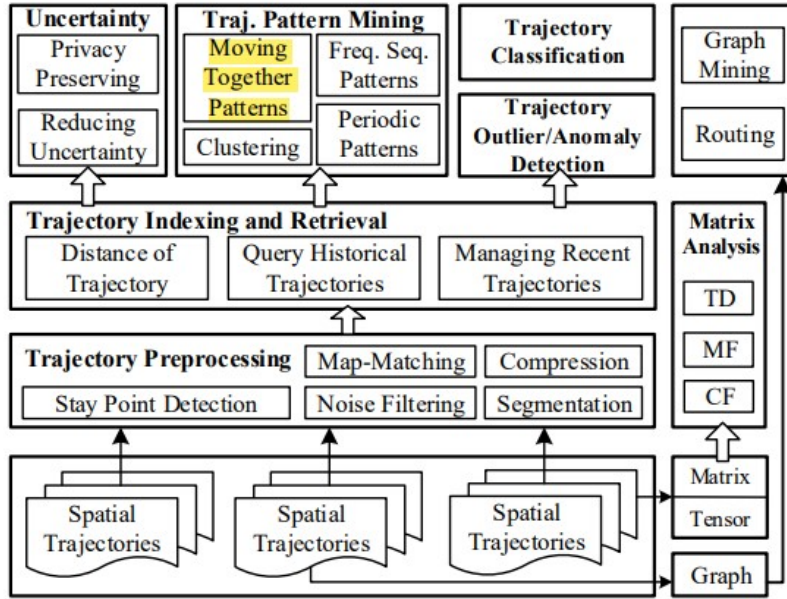


Figure 1 Paradigm of trajectory data mining

3.1 Obiettivo 2

Partendo dai punti di stop è possibile seguire due differenti approcci per calcolare il tempo totale di visita per ogni esibizione:

Approccio 1 Assegnare ad ogni stop point l'esibizione più vicina considerando il suo centroide, presnetata ambiguità per punti tra le intersezioni dei buffer delle exhibit

1. Creo buffer delle exhibit e calcolo il centroide

2. Assegno ogni stop point all'exhibit più vicina in base alla distanza minima tra stop point e centroide di ogni exhibit
3. Calcolo del visiting time per ogni exhibit

Approccio 2 Assegnare ogni stop point al centroide degli stop points dello stesso segment

1. Creo buffer delle exhibit e calcolo il centroide
2. Calcolo i centroidi del convex hull di ogni segmento di stop points
3. Considero solo i centroidi che intersecano il buffer delle exhibit e ne calcolo la distanza dal centroide di ogni exhibit
4. Filtro i risultati ottenuti mantenendo solo le coppie (C_{sps}, C_{exh}) che presentano la distanza minima
5. Assegno ad ogni segment l'esibizione più vicina, ottenendo quindi per ogni stop point l'esibizione a cui si riferisce
6. Calcolo del visiting time per ogni exhibit

3.2 Obiettivo 3

Per trovare il numero di persone che in un certo periodo temporale si trovano nei pressi della stessa esibizione è stato utilizzato MobilityDB (ref).

In particolare:

Approccio 1

1. Creo buffer delle exhibit e calcolo il centroide
2. Assegno ogni stop point all'exhibit più vicina in base alla distanza minima tra stop point e centroide di ogni exhibit
3. Tramite l'utilizzo della funzione tgeompoint_seq, dai singoli punti sono state create le traiettorie come tipo nativo di MobilityDB
4. Sono stati raggruppati i risultati per persona ed exhibit contando il numero di persone che in un certo periodo temporale si trovano nei pressi della stessa esibizione

Approccio 2

1. Creo buffer delle exhibit e calcolo il centroide
2. Calcolo i centroidi del convex hull di ogni segmento di stop points
3. Considero solo i centroidi che intersecano il buffer delle exhibit e ne calcolo la distanza dal centroide di ogni exhibit
4. Filtro i risultati ottenuti mantenendo solo le coppie (C_{sps}, C_{exh}) che presentano la distanza minima
5. Tramite l'utilizzo della funzione `tgeompoint_seq`, dai singoli punti sono state create le traiettorie come tipo nativo di MobilityDB
6. Sono stati raggruppati i risultati per persona ed exhibit contando il numero di persone che in un certo periodo temporale si trovano nei pressi della stessa esibizione