

Trajectory Clustering Algorithms - Two scalable and Comovement-based approaches.

Federico Naldini

Alma Mater Studiorum - Università di Bologna, Cesena.

federico.naldini3@studio.unibo.it

15/10/2019

- 1 Comovement patterns
- 2 GCMP: un framework generico per il mining di Comovements
 - Obiettivi e definizione del problema e dei parametri
 - Algoritmo di risoluzione
 - Generazione degli Snapshot
 - Star Partitioning
 - Apriori Enumeration
- 3 Scalable Distributed Subtrajectory Clustering

Che cosa sono i *Comovements patterns*?

I *Comovements patterns* sono raggruppamenti di oggetti che hanno viaggiato assieme per un certo periodo di tempo.

L'interesse per questi insieme può essere basato su diversi fattori, come ad esempio il numero degli elementi, la durata del viaggio, la loro effettiva vicinanza e il criterio utilizzato per calcolarla.

A seconda delle differenti caratteristiche utilizzate nel definire i raggruppamenti, è possibile definire delle tipologie di raggruppamenti.

Le tipologie di *Co-movements pattern* possono essere divise sulla base di diversi fattori; tra tutti spicca in particolare la metrica di similarità utilizzata. Possono essere impiegate due misure:

- **Similarità basata sulla densità**
- **Similarità basata sulla distanza**

- **Convoy**: Identifica raggruppamenti di oggetti che hanno percorso traiettorie simili per almeno T istanti consecutivi, negli algoritmi classici prima viene applicato un algoritmo di *clustering* basato sulla densità e successivamente viene effettuato *pruning* sulla base del criterio temporale sopraespresso.
- **Swarm**: Rispetto a **Convoy** scarta il vincolo di sequenzialità degli istanti temporali, basandosi solo sui vincoli spaziali.
- **Platoon**: Rimuove il vincolo descritto da **Convoy**, sostituendolo con un vincolo locale sugli istanti consecutivi tramite un parametro L che identifica la lunghezza minima di ogni sottosequenza di istanti consecutivi.
Inoltre aggiunge un altro parametro K , che identifica la lunghezza minima della sequenza temporale T di ogni raggruppamento.

- **Flock**: La stessa idea presentata in **Convoy**, ma utilizzando una metrica di similarità basata sulla definizione di uno spazio chiamato *disk* di raggio r .
- **Group**: Rilassa il vincolo temporale introdotto da **Flock** introducendo un vincolo sulla lunghezza delle sequenze di istanti consecutivi, analogamente a quanto fatto in *Platoon*; tuttavia a differenza di quest'ultimo non introduce un vincolo sulla lunghezza totale della sequenza.

Il framework **GCMP** si pone come obiettivo di mettere a disposizione degli utilizzatori una piattaforma configurabile per realizzare tutte le tipologie di *Co-movement mining* sopradescritte.

In particolare definisce diversi parametri per la definizione del problema:

General Co-movement Pattern Mining: Parametri 1

- **M**: Numero minimo di elementi presenti in un raggruppamento per considerarlo interessante.
- **K**: Numero minimo di istanti temporali in cui un certo raggruppamento esiste.

- **L**: Data una sequenza temporale di istanti T , si identificano z sottosequenze tali che ogni sottosequenza è composta da istanti consecutivi(ad esempio con $T = (1,2,3,5,6)$ si ottengono due sottosequenze $T' = (1,2,3)$ e $T'' = (5,6)$); **L** identifica la lunghezza minima accettabile di tutte le z sottosequenze così individuate(ad esempio con $L = 3$ la sequenza non rispetta il vincolo, mentre con $L = 2$ sì).

Una sequenza che rispetta il vincolo sopradescritto viene definita *L-consecutive*

- **G**: Data una sequenza temporale di istanti T , **G** identifica il massimo *skew* accettabile tra un elemento della sequenza e il successivo. Una sequenza si definisce *G-connected* se per ogni coppia di elementi consecutivi al suo interno lo *skew* in questione è minore o uguale a **G**

Un **GCPM** trova un set di oggetti O , rimasti assieme per una sequenza di istanti T che soddisfa i seguenti vincoli:

- *Closeness*: per ogni istante di T , gli oggetti O devono appartenere allo stesso *cluster*.
- *Significance*: La dimensione del raggruppamento deve essere maggiore di M .
- *Duration*: La dimensione di T deve essere maggiore di K .
- *Consecutiveness*: T è *L-consecutive*.
- *Connection*: T è *G-connected*.

General Co-movement Pattern Mining: Parametri 3

Configurando i vari parametri, posso ottenere le cinque tipologie di *Co-movements*

| Pattern | M | K | L | G | Clustering |
|----------------|---------|---------|---------|----------------|-------------------|
| Group | 2 | 1 | 2 | $ \mathbb{T} $ | disk |
| Flock | \cdot | \cdot | K | 1 | disk |
| Convoy | \cdot | \cdot | K | 1 | density |
| Swarm | \cdot | \cdot | 1 | $ \mathbb{T} $ | density |
| Platoon | \cdot | \cdot | \cdot | $ \mathbb{T} $ | density |

Trajectory Clustering Algorithms - Two scalable and Comovement-based approaches.

Federico Naldini

Alma Mater Studiorum - Università di Bologna, Cesena.

federico.naldini3@studio.unibo.it

15/10/2019