Visually–driven Analysis of Movement Data By Progressive Clustering

S. Rinzivillo
D. Pedreschi
KDD Lab, University of Pisa -- {rinziv,pedre}@di.unipi.it

M. Nanni
F. Giannotti
KDD Lab, ISTI – CNR, Pisa -- {mirco.nanni,fosca.giannotti}@isti.cnr.it

N. Andrienko
G. Andrienko
Fraunhofer Institute IAIS -- {gennady.andrienko,natalia.andrienko}@iais.fraunhofer.de
Problem

- Trajectories are complex spatio-temporal constructs
- Need for methods to assess the (dis)similarity between trajectories
- A single distance function is non suitable
  - It is difficult to build
  - It requires much time to compute
  - It is difficult to interpret the results
Idea

- Progressive clustering
  - Provide the analyst with a library of distance functions, each with a clear meaning
  - Step refined analysis through the successive application of several distance measures
    - Start with simple and efficient measures (common ends)
    - Refine the obtained clusters with more sophisticated functions
Similarity Functions

**Unique Distance Function**
- Computationally expensive
- Complex definition
- Complex indexing strategies
- Wastes time in analysing also the noise
- Generates many clusters
  - Hard to describe and interpret

**Several distance functions**
- Very efficient
- Simpler definitions (usually based on local observations)
- Simple indexing strategies
- Refinement of the relevant objects
- Stepwise refinement of clusters
Density Based Clustering

K-means

Density-based

cluster 1

cluster 2

cluster 3

cluster 4
Process Overview

- Simple and very efficient distance measure
- More selective and particular distance functions (or more restrictive parameters)
- Dataset
  - Clusters
    - Subclusters
  - Noise
    - Subclusters
    - Noise

Knowledge
Progressive Clustering - Example

Common Ends
Eps: 500
MinNbs: 10

Largest cluster
~3.6k Trajs
Progressive Clustering - Example

Common Ends
Eps: 500
MinNbs: 10

Other clusters
Progressive Clustering - Example

Common Ends
Eps: 500
MinNbs: 10

Focus on three interesting clusters
Progressive Clustering - Example

Common Ends
Eps: 500
MinNbs: 10

Choose one cluster
Progressive Clustering - Example

Common Ends
Eps: 500
MinNbs: 10
+
Route Similarity
Eps: 1000
MinNbs: 5

Routes from center to NW
Progressive Clustering - Example

- Common Ends
  - Eps: 500
  - MinNbs: 10
- Route Similarity
  - Eps: 1000
  - MinNbs: 5

Routes from center to NW
Progressive Clustering - Example

Common Ends
Eps: 500
MinNbs: 10

Route Similarity
Eps: 1000
MinNbs: 5

Routes from center to NW

Clustered by OPTICS with distance threshold = 2000.0 and minimum number of objects 5.
Distance function: Route similarity.
Future work

- Other clustering methods
  - Hierarchical vs Density-based
  - Dendrograms vs Reachability Plot
- In-memory computation issues
  - Exploit indexing strategies for neighborhood searches
  - Clustering by sample
    - Select a subset of the whole dataset and identify the clusters
    - Assign the other objects to one of the selected clusters
- Feature-based clustering
  - Eg. Distinct clusters with common behaviors: [work, shopping, home]