Consiglio Nazionale delle Ricerche

UNIVERSITÀ DI PISA

Project 5: Systematic comparison of EPR models

The student should develop a systematic comparison of Exploration and Preferential Return (EPR) models using scikit-mobility. In particular, the student should compare the following models provided in the library:

- SpatialEPR
  - see also Human Mobility Modelling: Exploration and Preferential Return Meet the Gravity Model
- DensityEPR
  - see also Human Mobility Modelling: Exploration and Preferential Return Meet the Gravity Model
- DITRAS
  - see also Data-driven generation of spatio-temporal routines in human mobility

The student should run each model for 5,000 agents for a period of two weeks, on at least four different geographic areas (e.g., cities, regions), and using squared, hexagonal (h3), voronoi, and an official division in areas such as neighborhoods or census cells. For the Voronoi tessellation, the student can select the dataset they prefer (e.g., OpenCellID).

For example, for the SpatialEPR model, the student should perform the following list of experiments (similarly for the other models):

Model	Geographic Area	Tessellation
SpatialEPR	area1	squared
SpatialEPR	area1	h3
SpatialEPR	area1	voronoi
SpatialEPR	area1	official division
…	…	…
SpatialEPR	area2	squared
SpatialEPR	area2	h3
…	…	…
SpatialEPR	area3	squared

SpatialEPR	area3	h3

The models' realism (i.e., the realism of the generated trajectories) should be compared in terms of:

1. distribution of jump length (M_1);
2. distribution of the radius of gyration (M_2);
3. distribution of uncorrelated entropy (M_3);
4. number of distinct visited locations (M_4);
5. number of visits per location (M_5);
6. location frequency (M_6).

Use the appropriate scikit-mobility measures to compute the above metrics.

For given an area (e.g., city, region), compute the distributions M_1, …, M_6 across the agents for each of the three models (SpatialEPR, DensityEPR, DITRAS). Given a distribution M_i ($i \in \{1, \ldots, 6\}$), compute a proper binning for the relative distribution and compare all models' distributions pair by pair using the Root Mean Squared Error (RMSE). See this paper (Section Results, Model comparison and validation).

NOTE: some of the models (e.g., DensityEPR, DITRAS) need information about location relevance, so define a proper relevance measure and find a proper dataset on which to estimate it. Moreover, DITRAS requires the training of a Markov chain on some trajectory dataset.