Trajectory Recovery API Documentation

1 Preliminaries, Dependencies

The `trajectory_recovery` Python module provides an interface for evaluating datasets on the trajectory recovery algorithm proposed by Xu et al. in [1]. This document contains the API documentation for the `TrajectoryRecovery` class. The dependencies of this module, excluding the standard library, are:

- numpy [2]
- pandas [3]
- matplotlib [4]
- scipy [5]
- geopy [6]
- tqdm [7]

For the sake of brevity, we use the following abbreviations:

- N: The number of trajectories in the dataset.
- M: The number of locations in the dataset.
- T: The number of time steps in the dataset.
- D: The number of time steps that can occur in 24 hours.

2 API Documentation

`TrajectoryRecovery()`, the constructor, expects **all** of the following arguments (unless otherwise specified) in the given order:

- aggregated_dataset: *pandas.DataFrame*

  The aggregated dataset with exactly N rows and M columns. Rows must appear in chronological order. The dataset must begin at 00:00. The order of columns (locations) from left to right is used for the below.

- grid: *dict*

  Location information that maps i, (the i-th location above) to a tuple representing its location in space. They may be mapped to cartesian coordinates, or given as latitude and longitude coordinates.

- num_trajectories: *int*

  N, the number of trajectories in the dataset.

- num_locations: *int*

  M, the number of locations in the dataset.

- num_timesteps: *int*

  T, the number of time steps in the dataset.

- num_intervals_per_day: *int*

  D, the number of time steps that can occur in 24 hours.

- cartesian: *bool*

  Whether the locations in the grid are mapped to cartesian coordinates, or are latitude and longitude coordinates. If this is not provided, then this is set to `True`.

`TrajectoryRecovery.run_algorithm()`

Runs the algorithm on the initialised aggregated dataset.

Returns *None*.

`TrajectoryRecovery.evaluate(truth_dataset)`

Evaluates the current predictions on a given truth dataset.

Returns a *dict* containing accuracy, recovery error, and top-k uniqueness metrics for the predicted and true datasets, for all $1 \leq k \leq 5$. It also contains a list of tuples where each (i, j) means that the i-th predicted trajectory was matched with the j-th true trajectory.

- truth_dataset: *list[list]*

  A 2D (N rows, T columns) list of true trajectories. The order of rows is not important, but each trajectory must contain the locations in chronological order. Each element is a tuple expressing the location.

`TrajectoryRecovery.visualise(timestep_range)`

Plots all the matched predicted and associated true trajectories within the given timestep range.

Returns a *list* of *matplotlib.pyplot* figures.

- timestep_range: *tuple[int, int]*

  If no time step range is given, then the range of $[0, \min(T, D))$ is used.

`TrajectoryRecovery.gain(trajectory_1, trajectory_2)`

Calculates the gain of two trajectories.

Returns a *float* of the calculated gain.

- trajectory_1 : *list[int]*
- trajectory_2 : *list[int]*

`TrajectoryRecovery.uniqueness(data, k)`

Calculates the top-k uniqueness of a dataset.

Returns a *float* of the calculated gain.

- data : *list[list]*

  A 2D (N rows, T columns) list of trajectories. Each trajectory is expressed as a list of sequential locations.

- k : *int*

`TrajectoryRecovery.get_predictions()`

Returns a 2D (N rows, T columns) list of the predicted trajectories.

`TrajectoryRecovery.get_results()`

Returns a *dict* containing the results of the most recent evaluation, including accuracy, recovery error, and top-k uniqueness metrics for the predicted and true datasets, for all $1 \leq k \leq 5$. It also contains a list of tuples where each (i, j) means that the i-th predicted trajectory was matched with the j-th true trajectory.

References

[1] F. Xu, Z. Tu, Y. Li, P. Zhang, X. Fu, and D. Jin, "Trajectory recovery from ash: User privacy is not preserved in aggregated mobility data," in *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, International World Wide Web Conferences Steering Committee, Apr. 2017.

[2] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, pp. 357–362, Sept. 2020.

[3] T. pandas development team, "pandas-dev/pandas: Pandas," Feb. 2020.

[4] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[5] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[6] GeoPy Contributors, "GeoPy: Python Geocoding Toolbox." `https://geopy.readthedocs.io/`.

[7] Casual Programmer's Incremental Developments, "tqdm: A Fast, Extensible Progress Bar for Python." `https://github.com/tqdm/tqdm`.