

Column Information:

The format of each record is: <t, lon, lat, obj_id, flag, subtraj_id, subtraj_type, traj_id, port_id, port_title>. The first three columns are original AIS data while the rest columns are produced through a data enrichment methodology. The description of each column as well as the data enrichment methodology are presented in the following paragraphs.

Original data

The original data come from the AIS (Automatic Identification System) network and contain position reports of vessels. The original columns contained in the original data are:

- t: the timestamp which the AIS position report was recorded at.
- lon: the longitude of the position the vessel was located at timestamp t.
- lat: the latitude of the position the vessel was located at timestamp t.

Data enrichment methodology

Step 1: Removing MMSI

For privacy issues, the MMSI (Maritime Mobile Service Identity) number of each vessel was replaced by a random number. This created one additional column:

- obj_id: a number which is used to uniquely identify a moving object (i.e. a vessel).

Step 2: Discovering stops

The data was processed according to the methodology described in [1] in order to discover the stops found in each vessel's movement history. This created the following additional columns:

- flag: a flag indicating among the following cases, depending on the characterization that the algorithm of [1] gave to the position: 0 → normal; 1 → stopped; 2 → low speed; 3 → gap in report; 4 → turn; 5 → speed changed.
- subtraj_id: a unique ID for the current part (sub-trajectory) of the vessel's movement.
- subtraj_type: a flag indicating among the following cases: 0 if the vessel is moving; 1 if the vessel is stopped outside a port; 2 if the vessel is stopped inside a port (the distinction between 1 and 2 is due to the fact that when the detected stop is in the area of a port, it is actually produced in the next step of the processing, called "Discovering trips between ports").

Step 3: Discovering trips between ports

After the stops were discovered (according to the previous step), the Wikimapia database of ports [2] was used in order to determine which of the stops are within a predefined area (circular area of radius = 1 n.m.) from a known port. This made it possible to infer, for each vessel, the trips between ports. The columns added by this step are:

- traj_id: A unique ID for a vessel's trip.
- port_id: The ID of the destination port (in accordance to the respective Wikimapia ID) of the trip.
- port_title: The name of the destination port (taken from Wikimapia) of the trip.

References:

- [1] K. Patroumpas, "Online Tracking and Summarization over Streaming Maritime Trajectories", In Proceedings of MOVE Workshop on Moving Objects at Sea, Brest, France, June 2013. Available online at:
<http://www.dbnet.ece.ntua.gr/pubs/details.php?id=1717&clang=1>.
- [2] Wikimapia, Category "harbor" with ID "754". Available online at:
<http://wikimapia.org/object/category/?id=754>

Last updated: March 17, 2014 by Marios Voudas.