Moving Objects Analytics: The Case of Maritime Analytics

Dimitris Zissis
Moving Objects Analytics ➔ Maritime Analytics

- Trajectory Data Mining (theory & practice)
  - Data Sources
  - trajectory pre-processing
  - Indexing and retrieval
  - trajectory pattern mining
Computing with spatial data

- A spatial trajectory is a trace generated by a moving object in geographical space, usually represented by a series of chronologically ordered points, e.g. \( p_1 \rightarrow p_2 \rightarrow \cdots \rightarrow p_n \), where each point consists of a geospatial coordinate set and a timestamp such as \( p=(x,y,t) \).

<table>
<thead>
<tr>
<th>ID</th>
<th>POS_ID,MMSI STATUS,STATION,SPEED,LAG,COURSE,HEADING,TIMESTAMP,SHIP_ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1094062019</td>
<td>667002119,0,1466,91,&quot;27,41997&quot;,&quot;36,84816&quot;,208,511,16/09/14 12:54,763967</td>
</tr>
<tr>
<td>1094076566</td>
<td>667002119,0,1466,91,&quot;27,41625&quot;,&quot;36,84376&quot;,218,511,16/09/14 12:56,763967</td>
</tr>
<tr>
<td>1094091446</td>
<td>667002119,0,1468,93,&quot;27,41197&quot;,&quot;36,83942&quot;,218,511,16/09/14 12:58,763967</td>
</tr>
<tr>
<td>1094108793</td>
<td>667002119,0,1468,93,&quot;27,40693&quot;,&quot;36,83435&quot;,219,511,16/09/14 13:01,763967</td>
</tr>
<tr>
<td>1094132136</td>
<td>667002119,0,1466,93,&quot;27,4004&quot;,&quot;36,82856&quot;,221,511,16/09/14 13:04,763967</td>
</tr>
<tr>
<td>1094147871</td>
<td>667002119,0,1466,95,&quot;27,39584&quot;,&quot;36,82366&quot;,212,511,16/09/14 13:06,763967</td>
</tr>
<tr>
<td>1094162364</td>
<td>667002119,0,1466,95,&quot;27,39234&quot;,&quot;36,81873&quot;,208,511,16/09/14 13:08,763967</td>
</tr>
<tr>
<td>1094176311</td>
<td>667002119,0,1466,94,&quot;27,38876&quot;,&quot;36,81381&quot;,211,511,16/09/14 13:11,763967</td>
</tr>
<tr>
<td>1094193923</td>
<td>667002119,0,1466,93,&quot;27,38432&quot;,&quot;36,80786&quot;,209,511,16/09/14 13:13,763967</td>
</tr>
<tr>
<td>1094211580</td>
<td>667002119,0,1466,94,&quot;27,38013&quot;,&quot;36,80176&quot;,208,511,16/09/14 13:16,763967</td>
</tr>
</tbody>
</table>
System Model for LBSs

- The locations of tracked moving objects are reported to the location server.
- The LBS applications submit queries to the server to retrieve moving object data for analysis or other needs.
Big (spatial) Data
Well...the data is BIG

• Volume
  • 70GB per day
    • 50 bytes per message

• Speed
  • 1 message per minute per vessel
  • >0.5 Billion messages per day
  • >200,000 streams

• Variety
  • Multiple receiving sources
  • Data fusion from various different sources
Data Sources
Maritime Big Data

• While in the past surveillance data was scarce, today there is a multitude of tracking systems both on board the vessel but also “above”
  • Collaborative Systems
    • Automatic Identification System (AIS)
    • Long-Range Identification and Tracking (LRIT)
    • Vessel Monitoring Systems (VMS)
  • Non collaborative systems
    • Cameras
    • Synthetic Aperture Radar (SAR)
Automatic Identification System

• The AIS is a **collaborative**, self-reporting system that allows marine vessels to broadcast their information to nearby vessels and on-ground base stations.

• It uses digital radio signals to exchange real time information between vessels and shore based stations on dedicated VHF frequencies.
  • Collision detection
  • Although mandatory for large commercial vessels to carry device, it is not mandatory to use it.
  • Not a replacement of radar as it cannot detect land masses, navigation beacons and vessels not equipped with AIS
**Class A Transceivers**

Class A AIS transceivers transmit and receive AIS signals. AIS transceivers are currently mandatory on all commercial vessels exceeding 300 tons that travel internationally (SOLAS vessels).

The following information can be transmitted by a Class A AIS system:

- **Static data.** Includes information such as vessel name, vessel type, MMSI number, call sign, IMO number, length, beam and GPS antenna location.
- **Voyage related data.** Includes information such as draft, cargo, destination, ETA and other relevant information.
- **Dynamic data.** Includes information such as time (UTC), ship’s position, COG, SOG, heading, rate of turn and navigational status.
- **Dynamic reports.** Ship’s speed and status.
- **Messages.** Alarms and safety messages.

Remember that not all vessels will transmit all of the information.

**Class B Transceivers**

Class B AIS transceivers transmit and receive AIS signals, but use a reduced set of data compared to Class A (see *Data Summary*). A Class B AIS transceiver can be fitted on any vessel not fitted with a Class A transceiver, but is not mandatory aboard any vessel.

<table>
<thead>
<tr>
<th>Data</th>
<th>Class A (receive)</th>
<th>Class B (send)</th>
<th>Class B (receive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Call sign</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>IMO number</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Length and beam</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Antenna location</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Draft</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Cargo Information</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Destination</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>ETA</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Time</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Ship’s position</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>COG</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SOG</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Gyro heading</td>
<td>Yes</td>
<td>Yes*</td>
<td>Yes</td>
</tr>
<tr>
<td>Rate of turn</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Navigational status</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Safety message</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
It is impossible to capture a lossless trajectory

• To capture though the accurate and complete trajectory of a moving object, is almost impossible in real conditions, due to the inherent limitations of data acquisition and storage mechanisms.

• Thus it can be captured as a time stamped series of location points denoted as \( p_0(x_0, y_0, t_0), p_0(x_1, y_1, t_1), \ldots, p_n(x_n, y_n, t_n) \), where \( x_i, y_i \) represents geographic coordinates of the moving object at time \( t_i \) and \( N \) is the total number of elements in the series. To generate the trajectory, a sensor needs to acquire its coordinates \( x, y \) at time \( t \).

\[
\sum \text{traj}_1 = p_0p_1, p_1p_2, p_2p_3, p_3p_4, p_4p_5, p_5p_6, p_6p_7, p_7p_8 \\
\sum \text{traj}_2 = p_0p_1, p_1p_3, p_3p_4, p_4p_7, p_7p_8
\]

\( \sum \text{traj}_1 \neq \sum \text{traj}_2 \)
Non uniform temporal and spatial data

- AIS information is classed as either static or dynamic. Static is broadcast when data has been amended or upon request or by default every 6 minutes.

<table>
<thead>
<tr>
<th>Class A systems</th>
<th>Reporting rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ships Dynamic Conditions</td>
<td></td>
</tr>
<tr>
<td>At anchor or moored</td>
<td>3 Minutes</td>
</tr>
<tr>
<td>0-14 knots</td>
<td>10 Seconds</td>
</tr>
<tr>
<td>0-14 knots and changing course</td>
<td>3⅓ Seconds</td>
</tr>
<tr>
<td>14-23 knots</td>
<td>6 Seconds</td>
</tr>
<tr>
<td>14-23 knots and changing course</td>
<td>2 seconds</td>
</tr>
<tr>
<td>Faster than 23 knots</td>
<td>2 seconds</td>
</tr>
<tr>
<td>Faster than 23 knots and changing course</td>
<td>2 seconds</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class B systems</th>
<th>Reporting rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ships Dynamic Conditions</td>
<td></td>
</tr>
<tr>
<td>0 to 2 knots</td>
<td>3 Minutes</td>
</tr>
<tr>
<td>Above 2 knots</td>
<td>30 Seconds</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other AIS sources</th>
<th>Reporting rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td></td>
</tr>
<tr>
<td>Search and Rescue (SAR) aircraft</td>
<td>10 seconds</td>
</tr>
<tr>
<td>Aids to navigation</td>
<td>3 minutes</td>
</tr>
<tr>
<td>AIS base station</td>
<td>10 seconds or 3.33 seconds, depending on operating param</td>
</tr>
</tbody>
</table>
Non uniform temporal and spatial data
Strait of Gibraltar. Ships (there are many) show up as bright dots.
ship targets seen next to the track through the sea ice.

Coastal Cameras
Trajectory Pre-processing
Stay Point Detection

- Spatial points are not equally important in a trajectory.
- Some points denote locations where objects have stayed for a while
  - such as shopping malls and tourist attractions, or gas stations where a vehicle was refueled.

With such stay points, we can turn a trajectory from a series of time-stamped spatial points $P$ into a sequence of meaningful places $S$,

$$P = p_1 \rightarrow p_2 \rightarrow \cdots \rightarrow p_n, \Rightarrow S = s_1 \Delta t_1 \rightarrow s_2 \Delta t_2 \rightarrow \ldots \Delta t_{n-1} \rightarrow s_n,$$

![Diagram of stay points in a trajectory](image-url)
Defining port boundaries

A distributed approach to estimating sea port operational regions from lots of AIS data
Scalable and distributed sea port operational areas estimation from AIS data

A distributed approach to estimating sea port operational regions from lots of AIS data
>57 million AIS messages! recorded during the month of March 2015

Scalable and distributed sea port operational areas estimation from AIS data
Compression

• Given a trajectory that consists of a full series of time-stamped points, a batched compression algorithm aims to generate an approximated trajectory by discarding some points with a negligible error from the original trajectory.

• A well-known algorithm, called Douglas-Peucker, is used to approximate the original trajectory.

• The idea of Douglas-Peucker is to replace the original trajectory by an approximate line segment, e.g. $p_1p_2$.

• If the replacement does not meet the specified error requirement (Perpendicular Euclidean Distance is used in this example), it recursively partitions the original problem into two sub-problems by selecting the point contributing the biggest error as the splitting point.
On-line Compression – Sliding Window

- Fit the location points in a growing sliding window with a valid line segment and continue to grow the sliding window until the approximation error exceeds some error bound.

1. First initialize the first location point of a trajectory as the anchor point \( p_a \) and then starts to grow the sliding window.

2. When a new location point \( p_i \) is added to the sliding window, the line segment \( p_a p_i \) is used to fit all the location points within the sliding window.

3. As long as the distance errors against the line segment \( p_a p_i \) are smaller than the user-specified error threshold, the sliding window continues to grow. Otherwise, the line segment \( p_a p_{i-1} \) is included as part of the approximated trajectory and \( p_i \) is set as the new anchor point.

4. The algorithm continues until all the location points in the original trajectory are visited.
Trajectory Indexing / Retrieval
Trajectory Data Management

- **Spatial Databases**
- **Queries**
  - Range queries
  - KNN queries
- **Distance metrics**
  - The distance between a point \( q \) and a trajectory
  - The Distance between two trajectories
  - The distance between two trajectory segments
- **Indexing structures**
- **Retrieval algorithms**
Uncertainty
Uncertain Trajectories

- As the location of a moving object is recorded at a certain time interval, the trajectory data we obtain is usually a sample of the object’s true movement. On one hand, the movement of an object between two consecutive sampling points becomes unknown (or called uncertain). To this end, we expect to reduce the uncertainty of a trajectory.

- On the other hand, in some applications, to protect a user’s privacy that could be leaked from her trajectories, we need to make a trajectory even more uncertain.
Interpolation or Gap filling
Traj. Pattern Mining
Trajectory pattern mining

Common Behaviour Detection: flocks, convoys, moving clusters, group patterns and swarms
Trajectory Clustering

• Route detection and definition
Why is this useful though?
Routes as traffic corridors (1)

• In the real world though, ships do not travel on these thin lines, as traffic corridors have a variable width, volume, and distribution

• identifying the specific characteristics of each corridor, and fundamentally the ship traffic dynamics.
Patterns of Life

Routes as traffic corridors (2)

Global Tanker PoL (more than 100 trips per route)

Tankers PoL in Mediterranean (with more than 100 trips)

Trajectory Classification/
Anomaly Detection
Detecting real world “anomalies”

Incident 1: The “Costa Concordia” grounding (13 January 2012)
Detecting real world “anomalies”

Incident 1: The “Costa Concordia” grounding (13 January 2012)
Graph Mining
Visualisation
The MarineTraffic AIS toolbox

An open-source tool developed from Marinetransfor Research

The MarineTraffic AIS toolbox provides a number of modules to support handling AIS data while improving their transformation into actionable visualisations such as density maps. The code is written in python for simplicity, readability, and overall ease of use. We look forward to seeing what people do with the MarineTraffic Toolbox and would like to hear your comments.

We will be releasing the software as a beta version over the next few weeks because we wanted to get feedback as soon as possible.

AIS Cleaning Module

Trajectories are never perfectly accurate due to sensor noise and other factors. In most situations it is necessary to apply algorithmic techniques to the data to smooth the noise and potentially decrease the error in the measurements.

This module also includes a number of simple data reduction techniques. The main objective of such trajectory reduction techniques is to reduce the size of the dataset so as to make it operable without compromising too much of its precision.

This module reads data as csv, it applies filters and delivers the cleaned data-set into the same csv format. The filters applied include removing empty fields, invalid movement data, invalid vessel details, special characters and downsampling the data according to user-defined parameters.

Density Maps Module

Density Maps support the improved understanding of vessel traffic, through providing a bird's eye view of vessel behavior either at a regional or global scale.

The term “vessel density” has several co-notations and thus is used with several meanings in this domain. Therefore, vessel density can refer to
The pandemic’s impact on global maritime mobility

The work is collaboration with several colleagues and friends from US and EU: Leonardo M. Millefiori, Paolo Braca, Stefano Marano, Peter K. Willett, Giannis Spiliopoulos, Sandro Carniel.
One of the most disruptive crises ever...

- To prevent the outbreak of COVID-19, many countries all around the world went into **full lockdown** in the first half of 2020.

- **Unprecedented containment measures**
  - Produced changes to all aspect of social life
  - Dramatically **changed mobility patterns**

- **Shipping industry** accounts alone for more than **80% of world trade**
  - Reduced ship mobility could imply reduced goods mobility on the global scale

- Need to **assess qualitatively and quantitatively** the impact of lockdowns on global shipping mobility

---


Data-driven approach to COVID-19 data

- Data-driven approach to COVID-19 data
  - **Global maritime mobility**
  - Epidemiological curve (upper plots)
    - Risk vs Detection Delay
    - Forecast evolution
- Unprecedented data analysis of the maritime mobility
  - Historical Automatic Identification System (AIS) dataset
    - 55 TB stored in a big-data architecture
    - 1 trillion AIS messages from more than 50,000 ships
    - Terrestrial network of over 4,500 receiving stations from 140 countries.
  - The processing is based on a distributed Apache Spark™ cluster of 40 virtual cores and 128 GB of RAM.

---


Monthly Cumulative Navigated Miles (CNM)

• Unprecedented decrease of the global mobility – CNM – after the lockdowns in 2020, compared with:
  • the previous years; and
  • the expected value in 2020 without pandemic

• Range of variation mobility [%]
  • Container $[-5.62, -13.77]$
  • Dry bulk $[+2.28, -3.32]$
  • Wet bulk $[-0.22, -9.27]$
  • Passenger $[-19.57, -42.77]$

Substantial increase in idle ships

- Substantial increase in idle ships across all types of ships/markets globally in the first 6 months of 2020
- Case of study
  - Daily active and idle ships in 2019 (in blue) and 2020 (in orange) of the supertankers, VLCC and ULCC.
  - Evident decrease (increase) of active (idle) ships from April, 2020, not present in 2019.
  - This is confirming the use of a significant subset of supertankers as oil storage.

Monthly CNM density difference between 2020 and 2019

- The considered time period is from 13 March to 13 April.
- Each grid cell is colored based on the variation of the 2020 value with respect to 2019.

Monthly CNM density difference – Suez Canal

Supply chain crisis in 2021-2022
Waiting times analysis

Work in progress...
Conclusion

- **Data-driven approach** to the global maritime mobility
  - Analysis based on the AIS
  - Unprecedented **decrease** of the mobility
  - However the supply chain **was not** disrupted showing a certain amount of **resilience**
  - More details at
  - [https://www.nature.com/articles/s41598-021-97461-7](https://www.nature.com/articles/s41598-021-97461-7)

Guide to Maritime Informatics

This unique volume formalises a body of knowledge for the emerging field of maritime informatics.

In the last 25 years, information systems have had a disruptive effect on society and business. Unlike recently though, the majority of passengers and goods were transported by sea in many ways similar to the way they were at the turn of the previous century. Gradually, advanced information technologies are being introduced, in an attempt to make shipping safer, greener, more efficient, and transparent. The emerging field of Maritime Informatics studies the application of information technology in maritime operations and activities.