Vessel Trajectories Outliers

Tomas Machado, Rui Maia, Pedro Santos and João Ferreira

May 24, 2018

Vessel Trajectories Outliers

Tomás Machado[1], Rui Maia[1], Pedro Santos[1], and Joao Ferreira[2]

[1] INOV, Lisbon, Portugal
[2] Instituto Universitário de Lisboa (ISCTE-IUL), Information Sciences, Technologies and Architecture Research Center (ISTAR-IUL), Portugal

Abstract. In this work we describe our first steps towards our H2020 project MARISA participation, where we intent to develop a tool-kit towards the identification of outliers in Vessel trajectories based on electronic data regarding position and time. These outliers can correspond to illegal activities that could be related with illegal immigration, drugs transshipment among others. We developed process tools that based on any electronic Vessel position systems, like Automatic Identification System (AIS) data, it is possible to extract routes in an unsupervised approach. At the same time identify non-conformities based on AIS data signal lost and to identify situation when two or more Vessels are approaching close to each other, called the rendezvous.

Keywords: Vessel, Trajectory, AIS, outliers, rendezvous

1 Introduction

Approximately 90% of the global trade being carried by the international shipping industry, turning the Ocean vital for the World's economy. Nowadays there are approximately 50,000 merchant ships trading internationally, and with the current demand this number tends to increase [3]. Although this efficient way of transportation presents threats that prevail in the maritime domain (i.e. piracy, trafficking of drugs, illegal immigration, arms proliferation, illegal fishing etc).

Automatic Identification System (AIS) is an automated tracking system, that broadcasts information via mobile maritime radio band aiding Vessels in collision avoidance. Imposed by the International Maritime Organization (IMO) every passengers ship must be equipped with an AIS device. Autonomously broadcast AIS messages contain Vessels kinematic information (including ship location, speed, heading, etc.) and static information (including ship name, ship Maritime Mobile Service Identity (MMSI), etc.), which can be transformed into useful information for maritime traffic manipulations (e.g. Vessel path prediction and collision avoidance). The introduction of AIS in the maritime domain, increased the volume of Vessel trajectory data exponentially, making human analysis and evaluation of this data extremely inefficient. Therefore, new effective ways of automatically mining this data show a great contribution for the future

[3] ICS Shipping and World Trade, www.ics-shipping.org/shipping-facts

of nautical surveillance. However, mining maritime trajectory data present several challenges, such as: 1) Maritime trajectory data possess the data uncertainty typical of moving object trajectories. Geo-referenced locations of trajectory positioned by location sensing techniques may be collected with spatial uncertainty due to computational error and signal loss or degradation associated with the positioning device, [1]; 2) maritime traffic is not constraint to roads; Vessels are free to navigate in open waters, increasing the complexity of anomaly detection drastically.

Vessel Anomalous Behavior can be subdivided into: 1) Kinematic behaviors relate to the motion of ships including the routes taken and speed of travel; 2) AIS transmission behaviors include the switching on or from AIS systems and changing the Vessels name or other details, [2].

Current work is under an H2020 project MARISA - Maritime Integrated Surveillance Awareness [4], where is created a tool-kit that provides a suite of services to correlate and fuse various heterogeneous and homogeneous data and information from different sources, including Internet and social networks. In the context of this project current work provides an analysis and synthesis of the traffic spatio-temporal data streams provided by the AIS cooperative self-reporting system requires a suitable degree of automation and efficiency to detect and characterize inconsistencies, anomalies, ambiguities and ultimately transform this information into usable and actionable knowledge. It provides the activity at sea as contextual information and patterns of life, referred as maritime routes and summarizes the maritime traffic over a given period of time and a given area.

2 State of Art

Vessel Behavior is as considered as a baseline in which abnormal behavior can be found. This baseline can occur as normal trajectories are various and constant, producing a normalcy model of Vessels dynamics, that Machine Learning Techniques can learn. A vast number of frameworks in which Vessel behaviour analyses with the purpose of anomaly detection are fully defined as integrated systems. The authors in [3], suggested the framework MT-MAD (Maritime Trajectory Modeling and Anomaly Detection), in which a given set of moving objects, the most frequent movement behaviour are explored, evaluating a level of suspicion hence detecting anomalous behaviour.

TREAD (Traffic Route Extraction and Anomaly Detection) is a framework in which an Unsupervised Route Extraction is used to create a statistical model of maritime traffic from AIS messages, in order to detect low-likelihood behaviours and predict Vessels future positions, [4].

A framework focused on Vessel interaction and Rendezvous, is proposed by [5], which uses a logically connected 3-phase process, reducing the volume of data that is processed by the sub-sequential phases, therefore prioritizing critical scenarios, that request human intervention.

[4] Marisa Project - www.marisaproject.eu

A Partition-and-Detect framework, in which trajectories are partitioned into a two-level of granularity, achieving high efficiency or high quality trajectory partitions is proposed in, [1]. Accessing the performance of these frameworks, is an arduous task, as there is no defined benchmark labelled sets where test assessment can be performed, [6]. Although a solution for constructing an AIS database, with the potential value for being used as benchmark database for maritime trajectory learning, is proposed in [7].

As the volume of positional AIS data exponentially increases, it's important to find methods in which raw trajectories data can generate knowledge.

Trajectory learning is the process of learning motion patterns from trajectory data using unsupervised techniques, mainly clustering algorithms [8]. Morris and Trivedi [9], further categorize trajectory learning as a three-step procedure: 1) Trajectory Preprocessing; 2) Trajectory Clustering; and 3) Path Modelling. Pallotta proposed a method that enriches the raw Vessels trajectories with a description of the ship movements, labeling the raw trajectories as 'Stationary' or 'Sailing' [4].

A way to discretize a trajectory is to represent a trajectory into a spatial grid in which a cell represents a geographical area with a defined size. Analysisng this grid is allows a effective discover of frequent regions, [3]. The authors in, [9], propose a method to dynamically analysing a trajectory, with the emphasis on the learning of AP (Activity Path) and the discovery of POIs (Point of Interest), which can indicate common Vessel destinations (e.g. frequent fishing zones, ports, etc.).

3 Methodology

Figure 1, represents our development defined as a process, towards the contributions specified above. Our process is partitioned into 3 major parts, Data preprocessing, Route extraction and Anomaly Detection

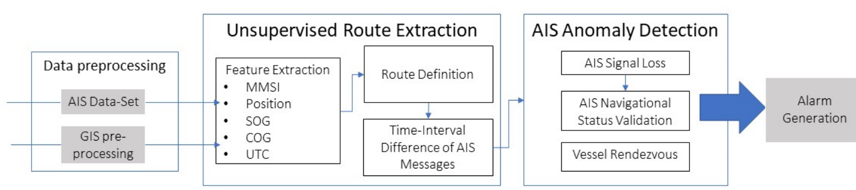

Fig. 1. Implemented Process towards our goals of Vessel Rendezvous and AIS signal loss.

3.1 Data Preprocessing

The open-source AIS data type file is a raw database file(.dbf), that permits low efficiency, data manipulations. Thus, it becomes important to transform this data format into a more efficient Data Wrangling format. By the means of an open-source Geographic Information System[5], we transformed the raw data into csv format. As dbf to csv transformations are time consuming, and a decent sized data-set is achieved with just one area. The chosen area 10, represents data whose longitude is from -120 to -126 and latitude is from 30 to 50 .

3.2 Unsupervised Route Extraction

We developed a efficient method for Unsupervised Route Extraction, based on [4], work. Our method can be fed with either a AIS Data Stream, an AIS Data Base, or Unprocessed AIS Data Base.

Feature Extraction By extracting only the relevant features, related to the Vessels Kinematics, and Navigational Status, we are able to reduce significantly the amount of data that we processed. Thus, increasing the effectiveness of our Anomaly Detection methods.

Route Definition A simple, but not as effective way to represent a trajectory, as more effective ways are proposed in the literature, is to represent a trajectory with no compression. This can be easily achievable as Vessel are identified with unique ID, and are obliged to broadcast their AIS information in semi-continuous rates. Thus, we consider a Vessel Trajectory as the aggregation of all its broadcasts, defined as a set of multidimensional-points represented as:

$$TR_{MMSI} = p1, p2, p3, p4, \cdots, pn \tag{1}$$

Where each multidimensional point p is defined as:

$$p = [t, x, y, SoG, CoG] \tag{2}$$

Time-Interval AIS Broadcast Our, definition of Trajectory, is further enriched by determining the difference between every t, t_{-1}, transmission, this allows an efficient way to manipulate data, so AIS Signal Loss can be found.

3.3 AIS Anomaly Detection

Vessel anomaly detection is a research field which poses an immense level of complexity. We constraint our work, to some Vessel anomalies, which are described as requirements in the MARISA project.

AIS Navigational Status Validation AIS Navigational Status, describes the Vessel periodic navigational activity, according a set of static Status such as: under-way by engines, at anchor, moored, aground, engaged in fishing, under-way by sail, etc. A detailed list and description all this status is found in [6].

[5] QGIS Geographic Information System, http://qgis.osgeo.org
[6] Solas Chapter V Annex 17 AIS - www.mcanet.mcga.gov.uk

AIS Signal Loss Blocking AIS is a process that Vessels responsible use, to hide their position, as an anti-piracy defense, or possibly for illegal activities. This is done by simply turning off AIS equipment or by block antenna signal. We developed an real-time Heuristic based process that generates alarms, for Vessels that do not broadcast AIS information, for a period longer, than a set threshold.

Vessel Rendezvous A requirement imposed by the MARISA project was the development of services, able to detect and generate alarms when two or more Vessels are approaching close to each other. In the maritime world this can be considered as a possible anomaly, which is called rendezvous. The concept of rendezvous in the Maritime world is quite complex, as there are numerous legislation. For this work, and because the emphasis is on the alarm generation, a simplification of this Vessel interaction is assumed. Thus, Vessel Rendezvous is considered as the interception or closeness of two or more Vessels, in a configurable Time Period.

3.4 AIS Data-Set

The sources of AIS data our work, are derivative from two different types: 1) Public sources this includes data from Australia[7] and USA[8]; and 2) Confidential AIS and S-AIS, data sources stemming from Military Forces via the MARISA project. From the USA open-source data, we created a Data-Set of 1659 different Vessels, representing approximately 12,3 Million AIS broadcast, representative of the year 2017. Although, for the Experiments conducted in Section X, we have used a small subset composed of 38 Fishing Vessels, which represents approximately 280.000 AIS messages.

Table 1. Example of AIS data transmitted by Vessel, MMSI: 636081210; first 4 transmissions from a total of 6256 transmissions.

MMSI	X	Y	SOG	COG	Time
636081210	-125.993218	48.355773	14.3	73.300003	2014-02-27 13:33:02
636081210	-125.985303	48.357340	14.5	73.800003	2014-02-27 13:34:23
636081210	-125.979437	48.358500	14.6	73.099998	2014-02-27 13:35:23
636081210	-125.973353	48.359692	14.7	73.000000	2014-02-27 13:36:25

4 Experience

AIS Navigational Status Validation In our experiment, we discovered that a large number of Vessels do not keep their Navigational Status updated. A plausible cause; as the Navigational Status is manually introduced and updated, it can lead to expected Human error. Nevertheless, the wrong use of a Navigational

[7] Vessel Tracking Data, www.amsa.gov.au/Spatial/DataServices/DigitalData
[8] MarineCadastre Vessel Traffic Data, www.marinecadastre.gov/ais

Status can lead to fines by the Maritime Authorities, thus being considered an Anomaly.

Table 2. 32 Vessel subset AIS Navigational Status description.

AIS Navigational Status (Number - Name)	Count	%
0 - Under way using engine	150465	54
15 - Default	97264	35
7 - Engaged at Fishing	24444	9
5 - Moored	5462	2
8 - Under way Sailing	2138	0,7

By determining the actual movement characteristics of a Vessel, we are able to deduce if either a Vessels is Stopped or Moving. This simple extrapolation permits the validation, of most used Navigational Status, we further our Stopped/Moving method with the detection of AIS Signal Loss. In this work we do not Validate if a Vessel is Engaged at Fishing, as it presents a certain level of uncertainty, and cannot be solved with a simple Stopped / Moving validation, a possible solution for this case is presented by the authors in [10].

Vessel Rendezvous Algorithm 1, describes the implemented process in a algorithm way, using as input distance D_{tresh}, and a interval time-group T_g. For every single Vessel Track, the Track is partitioned into defined time-groups (e.g. a T_g of 5min). If two or more Vessels are on the same T_g, the distance between these Vessels is calculated, using the Haversine, (3). If this distance is less than the D_{tresh}, an alarm is generated for these two Vessels.

Algorithm 1: Vessel Rendezvous Method

Input : A set of defined AIS routes, S_{tr}; Time-Group, T_g; and a Distance threshold, D_{tresh};

Output: A set of two or more Vessels with a distance in NM and Time, $RES : [MMSI_{set}; Timestamp; D_{NM}]$;

Initiation: Partition S_{tr}, into sub-groups SG_{Ttr} ,defined by T_g.
foreach SG_{Ttr} *in* S_{tr} **do**
    if SG_{Ttr} *contains more than 2 routes* **then**
        Calculate HarversineDistance in NM over all $(SG_{Ttr})C2$; **if** *HarversineDistance* $> D_{tresh}$ **then**
            $RES = [MMSI_{set}; Timestamp; D_{NM}]$
        end
    end
end

Haversine Distance is used, as latitude and longitude features are represented in a spherical coordinate system, using on the following equation, as d represents the distance between the 2 points:

$$d = 2rsin^{-1}(\sqrt{sin^2(\frac{lat_2 - lat_1}{2}) + cos(lat_1)cos(lat_2)sin^2(\frac{long_2 - long_1}{2})}) \quad (3)$$

5 Results

Results evaluation were based on two developed process: 1) AIS Navigational Status Validation; and 2) Vessel Rendezvous.

From AIS Navigational Status validation experience, we were able to deduct that Vessels tend to not update their Navigational Status. Thus, by labeling the AIS data, with Stopped and Moving labels, and we are able to determine how many mislabeled Broadcast were sent.

Table 3. Results for AIS Navigational Status and AIS Signal Loss occurrence count, with 32 Vessel Subset

Broadcast AIS Status	Total Count	Moving Count	Stopped Count	Error Count - %	Error %
0 - Under way using engine	150465	54734	95731	95731	64
5 - Moored	5462	1038	4424	1038	19
8 - Under way Sailing	2138	1118	1020	1020	47
15 - Default	97264	57616	39541	*	*
AIS Signal Loss	411	-	-	-	-

The accuracy for the 15 - Default Navigational, cannot be measure as the other Statuses, this Status represents that Vessels have kept the default AIS Status for the whole Trajectory. Thus, our results presents the number of AIS Default Status, for Moving and Stopped kinematics characteristics.

Vessel Rendezvous test was conducted using a 2-minute time group, and 1,5 Nautical Miles as distance Threshold. Table 4, represents the collection of alarms generated, after conducting the Vessel Rendezvous test in the 32 routes sub-set. The table shows 10 occurrences, ordered by distance, from 4 different Vessels. Multiple occurrences can occur from the same two Vessel routes if these Vessels were close to each other more than the defined time-group parameter.

Table 4. Results obtained for Vessel Rendezvous, with the 32 Vessel subset.

TimeSlot	MSSI A	MMSI B	Distance in NM
2014-02-15 11:32:00	357103080	370000802	1.012450
2014-02-01 21:30:00	413104010	432000385	1.160439
2014-02-01 21:36:00	413104010	432000385	1.160971

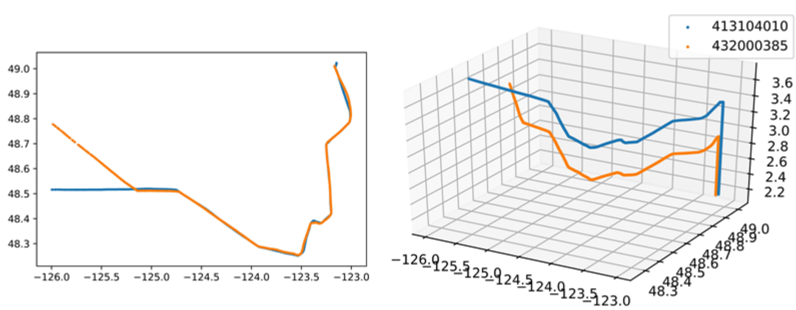

Fig. 2. Routes of MMSI: 413104010 and 432000385; axes representing (lat.,long.)[Left] and (lat.,long.,time)[Right]

In Fig. 4, the two Vessels, MMSI 413104010 and 432000385, are represented. These Vessels represent the Vessels that were close to each other, for the longest time-period, resultant from the Vessel Rendezvous test. Although, while Vessels are able navigate freely in open waters, it is common that certain routes are shared by Vessels, which can be caused due Maritime traffic. Which could, make normal the fact the Vessels were close to each other for the whole route. Although what could be considered abnormal, is the fact that when the Vessels are the closest to each other's, is the moment in which the Vessels change directions, which is shown in Fig.4(Right).

6 Conclusion

In the current research work, we developed tools to handle Vessel electronic data, mainly oriented to AIS and extract Vessel routes. From these routes, missing positions were identified and rendezvous situations. These processes generate alerts towards responsible control entities with the mission of checking these possible non-conformities in real-time. Until present moment, we are only able to validate Navigational Scenarios based on Stopped or Moving kinematics, the next step is the identification of fishing and constrained by her draught real scenarios, and further validation with MARISA marine partners.

References

1. Lee, Jae-Gil, et al. "Trajectory Outlier Detection: A Partition-and-Detect Framework." 2008 IEEE 24th International Conference on Data Engineering, 2008.
2. Lane, R O, et al. "Maritime Anomaly Detection and Threat Assessment." 2010 13th International Conference on Information Fusion, 2010.
3. Lei, Po-Ruey. "A Framework for Anomaly Detection in Maritime Trajectory Behavior." Knowledge and Information Systems, vol. 47, no. 1, 2015.
4. Pallotta, Giuliana, et al. "Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction." Entropy, vol. 15, no. 12, 2013.
5. Shahir, Hamed Yaghoubi, et al. "Maritime Situation Analysis: A Multi-Vessel Interaction and Anomaly Detection Framework." 2014 IEEE Joint Intelligence and Security Informatics Conference, 2014.
6. R. Laxhammar, "Anomaly detection for sea surveillance," 2008 11th International Conference on Information Fusion, 2008.
7. Mao, Shangbo, et al. "An Automatic Identification System (AIS) Database for Maritime Trajectory Prediction and Data Mining." Proceedings in Adaptation, Learning and Optimization Proceedings of ELM-2016, 2017.
8. N. Le Guillarme and X. Lerouvreur, "Unsupervised extraction of knowledge from S-AIS data for maritime situational awareness," Proceedings of the 16th International Conference on Information Fusion, 2013.
9. Morris, B.t., and M.m. Trivedi. "A Survey of Vision-Based Trajectory Learning and Analysis for Surveillance." IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, no. 8, 2008.
10. Natale, Fabrizio, et al. "Mapping Fishing Effort through AIS Data." Plos One, vol. 10, no. 6, 2015.