Big data for Maritime Domain Awareness: An AIS case study

Waldo Kleynhans

IMIS Global Limited, Fareham, UK

1 Introduction

The world's oceans is of critical importance to humanity as it is key to fisheries, shipping as well as the environment. From an economic perspective, it is estimated that 90% of all global goods and energy transportation are done by sea with millions of people being dependent on maritime related activities for their livelihood. As maritime activities increase globally, there exist a greater dependency on technology in the monitoring, control and surveillance of vessel activities. One of the most prominent systems for monitoring vessel activity is Automatic Identification System (AIS). AIS operates in the VHF band and transmits messages from vessels which can be received by other vessels, terrestrial shore stations as well as satellites.

When dealing with AIS data, two pertinent factors to conciser are:

1. **AIS data fidelity**: Due to the fact that AIS is broadcast in a non-secure channel, information could be manipulated / corrupted (such as malicious or inadvertently introduced false GPS positions and errors in vessel parameters). In addition, AIS receivers are not controlled in the same manner as AIS transmitters, which could introduce additional errors at the receiver side [1].
2. **Significant volume increase of AIS messages**: Due to the global increase in vessels fitted with AIS transmitters as well as the proliferation of satellite and terrestrial receiving stations there has been a significant increase in AIS messages received globally (estimated at over a 40% increase over the last four years [1]). While this increase in AIS data volumes is beneficial as this enriches the information available to maritime authorities, processing and storage of these large data volumes can become problematic especially when performing analytics based on historic vessel temporal and positional data.

By using advanced filtering and analytics, IMIS Global Limited has been able to process the AIS data stream to eliminate a large portion of the faulty data (as described in point 1) and is also focusing on the efficient storage and compression of vessel track history derived from AIS positional report data in an effort to deal with the challenges raised in point 2. Storing only fundamental data required to accurately construct vessel historic track (trajectory) data is also one of the key objectives of the datAcron initiative [2].

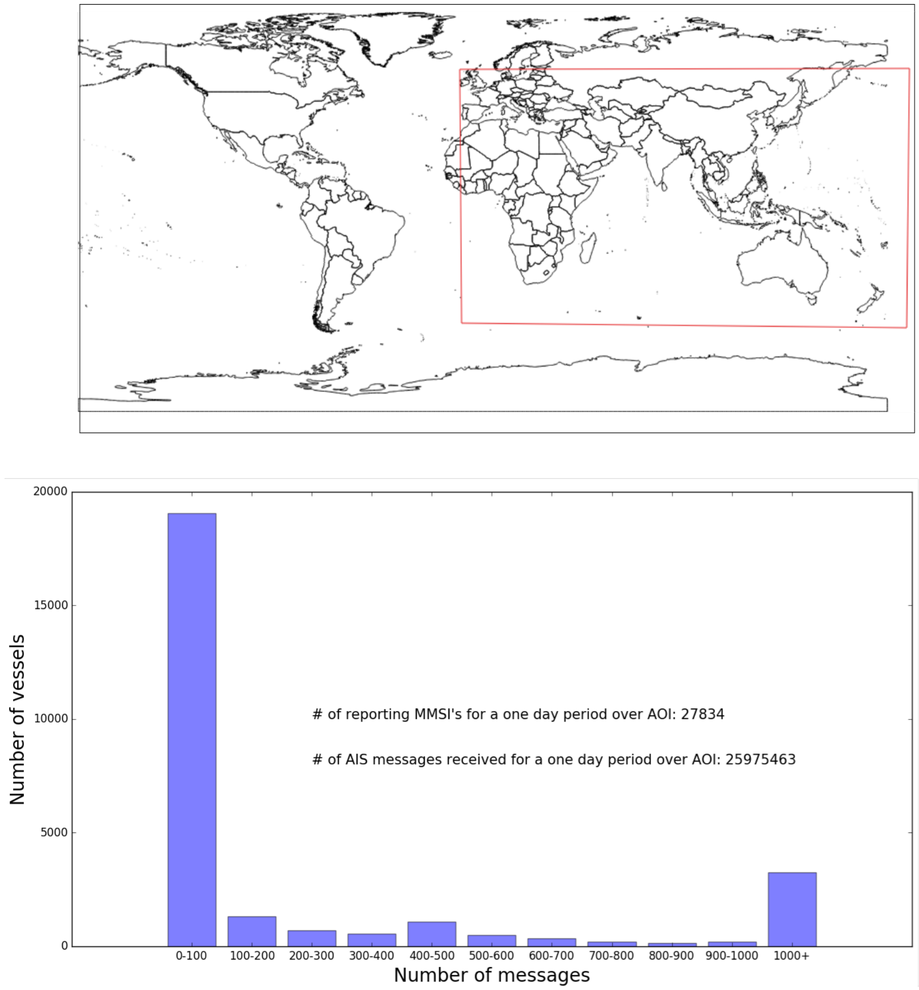

Fig. 1. Area of interest (top), received AIS data from this area is provided by more than a hundred data sources and includes terrestrial receivers, base stations as well as satellites. The distribution of the number of vessels vs. the number of received AIS messages is also provided (bottom)

2 Study Area

The study area as well as the typical data volumes for a 24 hour period for 17 March 2018 is represented in figure 1 (top) where it can be seen that the area of interest (AOI) covers a significant portion of the globe. AIS data from this area is provided by more than a hundred AIS data sources which includes terrestrial receivers, base stations as well as satellites. The total number of vessels that was received during the day was 27834. These vessels transmitted a combined total

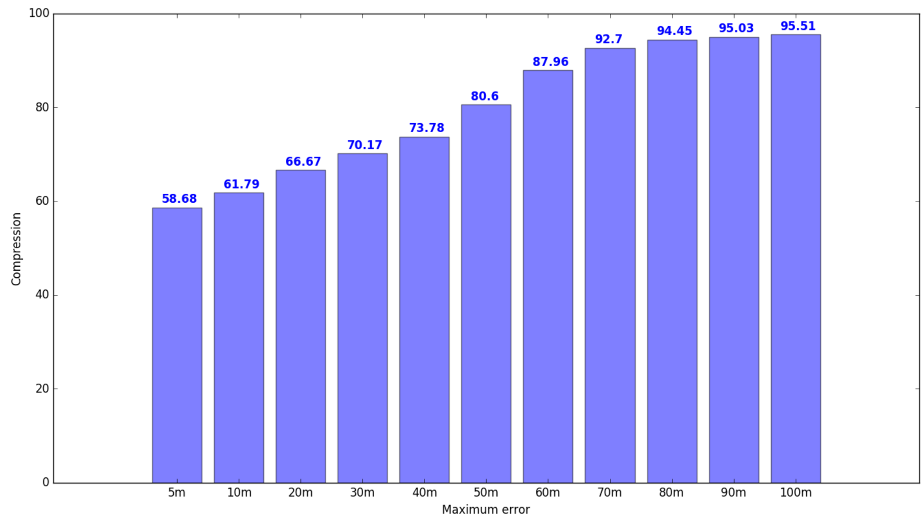

Fig. 2. Data compression as a function of the maximum distance error when when compared to the complete raw AIS positional dataset

of $\approx$ 26 Million AIS messages during the 24 hour period. The distribution of the typical number of messages received per vessel is also shown in figure 1 where it can be seen that the majority (>70%) of vessel's transmitted AIS messages were received only 0-100 times throughout the AIS network with ($\approx$ 15%) of vessel's total received AIS messages for the day being more than 1000. This can be attributed to the fact that AIS messages transmitted from vessels in the reception range of terrestrial AIS receivers are received far more frequently than from satellite AIS sensors whereas satellite AIS sensors, on the other hand, are able to cover very large areas and consequently can receive AIS message from significantly more vessels but at the much lower revisit rate. It was also found that 10% of the vessels generated >90% of the AIS messages received through the network of receivers in the AOI.

3 Results and discussion

From section 2 it is clear that a diverse AIS reception network generates an enormous amount of data. In many cases, especially when vessels are traveling at very low speeds or are stationary, and vessel track data are of interest, only a fraction of these reported messages are required for reconstructing an accurate vessel track when compared to using the entire compliment of raw AIS messages. An on-line lossy track compression methodology was used in the filtering process of the positional AIS data. In essence the method tracked the AIS message stream originating from each ship in an on-line fashion and compared the spatio-temporal information contained in each newly presented AIS message with that of the current vessel track history, by calculating a distance metric related to the maximum allowable spatio-temporal distance error from the current track, a decision was made to include the newly presented positional information or discard it. The goal was to be able to then do historic vessel track reconstruction as a function of the maximum allowable track distance error when compared to the complete raw AIS positional dataset. The results are shown in figure 2, it can be seen that, as expected, the compression ratio increases as the maximum distance error metric is relaxed. It was shown that a compression ratio of ($\approx 90\%$) could be achieved when the error metric of between 60m and 70m is selected.

References

1. Batty E. (2018) Data Analytics Enables Advanced AIS Applications. In: Doulkeridis C., Vouros G., Qu Q., Wang S. (eds) "Mobility Analytics for Spatio-Temporal and Social Data" MATES 2017. Lecture Notes in Computer Science, vol 10731
2. Georgios M. Santipantakis et. al.2017. "Specification of Semantic Trajectories Supporting Data Transformations for Analytics: The datAcron Ontology." In Proceedings of Semantics2017, Amsterdam, Netherlands, September 1114, 2017, 8 pages.