



Predicting the Time of Arrival at Port for Maritime Surface Ships in the Baltic Sea Using Recurrent Neural Networks

Jonatan Lahtivuori

Abstract

Here is where I would say what is in this document.

Contents

1	Introduction	1
2	Maritime vessels and traffic	2
2.1	Automatic identification system	2
2.1.1	Estimated Time of Arrival	4
2.2	HELCOM dataset	5
2.2.1	Description of HELCOM dataset features	6
2.2.2	Statistical information	6
2.3	Port of Naantali	7
3	Recurrent Neural Networks	8
3.1	Long Short Term Memory	8
4	Implementation	9
4.1	Libraries used	9
4.2	Data pre processing	9
4.2.1	Algorithm for extracting routes from dataset	9
4.2.2	Coordinate accuracy	10
4.2.3	Feature selection	10
4.2.4	Time series data preparation and cleanliness	10
4.3	ML model	10
4.4	Comparison model	11
5	Results	12
5.1	ETA prediction	12
5.2	Navigation in the archipelago	12
6	Discussion	13
7	Conclusion	14

1 Introduction

Describe the need for accurate ETA predictions for maritime traffic and what data is available what has been done already and what the thesis will contribute.

Structure of the thesis.

2 Maritime vessels and traffic

Vessels navigating the Baltic Sea

Importance of maritime traffic globally and timely operation

2.1 Automatic identification system

Automatic identification system (*AIS*) was introduced by the International Maritime Organization (*IMO*) in the early 2000's in accordance with the Safety of Life at Sea (*SOLAS*) treaty as an open communication tool for all maritime traffic. The main objectives of the treaty is to improve the safety of life at sea, protection of the marine environment and safety and efficiency of navigation [1]. In its general operating form, AIS operates on two VHF channels and all vessels or base stations within the vessels transmission range receive the messages transmitted and vice versa the vessel receives all messages broadcasted in its receiving range. The broadcasting is based on Self-organized Time Division Multiple Access ((*S*)*TDMA*) with a minimum of 2000 time slots per minute broadcasting capacity rate and the ship-to-ship communication for closer vessels takes precedence over vessels farther away which allows for sharing time slots, and thus overloading the available time slots [1].

The IMO requires in accordance with regulation V/19 of SOLAS, that all vessels with a gross tonnage of 300 or more on international voyages, cargo vessels of 500 gross tonnage or more not on international voyages and all passenger vessel disregarding the gross tonnage to be equipped with an AIS. Further, the EU requires new-built fishing vessels longer than 15 meters to be fitted with an AIS from November 2010 and existing vessels to install an AIS May 2014 at the latest [2]. There are two types of AIS Class A and Class B. Class A transceiver are the more common and are compliant with the IMO regulations whereas Class B transceivers are not subject to the full IMO AIS requirements. Class B transceivers are typically simpler, of lower cost and installed on smaller vessels of one's own choosing for improving the situational awareness. Class A transceivers take precedence over Class B transceivers when broadcasting.

The vessels onboard sensors and positioning systems interface with the AIS to allow for communicating the status of the vessel, both static and dynamic information. AIS also allows for voyage related information and safety-related information to be transmitted via the system, see Table 1. The AIS guidelines requires that there is a minimum display and keyboard for input and receive of data, for updating information manually and retrieving data received via AIS.

The information received by the AIS does not provide the full situational picture of the vessels surroundings and should only be used as a navigational and situational awareness aid.

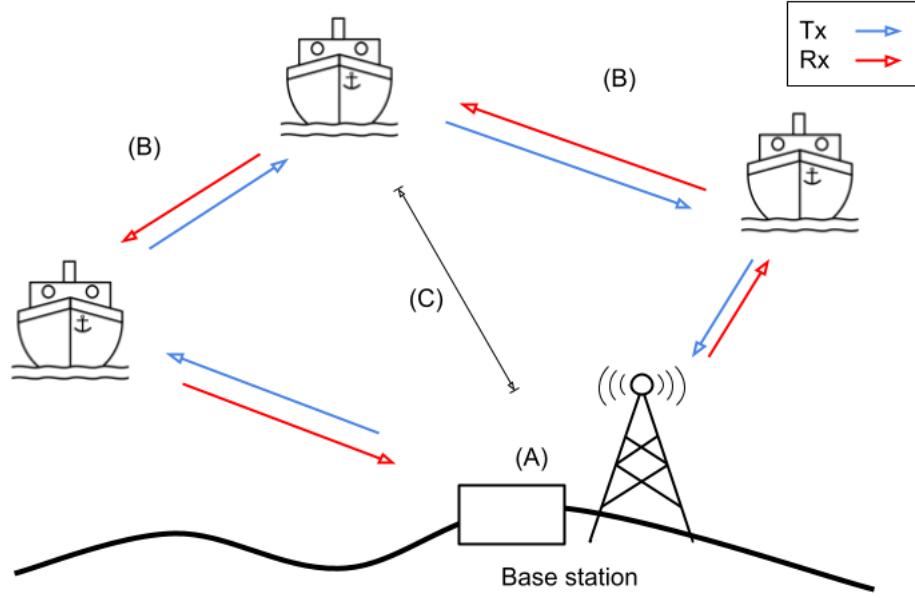


Figure 1: (A) Base station transmits information about other vessels, port data and possible hazards. Base station receives information from all vessels in range. (B) Ship-to-ship communication. Vessel broadcasts its current state (identifier, speed over ground, course over ground etc.) and receives data from all vessels in range, only limited by the number of possible time slots available. (C) Vessel outside range of transmitting AIS data to the base station still transmits information to other vessels.

The AIS integrated checks for data integrity, built-in integrity test (*BIIT*), is not capable of validating data, for example a non-functional sensor will report as not available to the AIS.

The information available to transmit over AIS is divided into three groups static, dynamic and voyage-related.

Data	Description
Static	
Maritime Mobile Service Identity (MMSI)	Set on installation, can change during vessels operational lifespan
Call sign and name	Set on installation, can change during vessels operational lifespan
IMO number	Set on installation
Length and beam	Set on installation or if changed
Type of ship	Selected from list of predefined values
Location of electronic positioning system	Set on installation, can be changed
Dynamic	
Ships position with accuracy indication and integrity status	Automatically updated from the position sensor
Position time stamp in UTC	Automatically updated from the main position sensor
Course over ground (COG)	Automatically updated from the main position sensor if available
Speed over ground (SOG)	Automatically updated from the main position sensor
Heading	Automatically updated from the vessels heading sensor
Navigational status	Manually entered by the OOW as needed
Rate of turn (ROT)	Automatically updated from the vessels ROT sensor or from the vessels gyro
Voyage-related	
Draught	Manually entered at the start of the voyage
Hazardous cargo (type)	Manually entered at the start of the voyage confirming the presence of such cargo
Destination and ETA	Manually entered at the start of the voyage and updates as needed
Route plan (waypoints)	Manually entered at the start of the voyage
Safety-related	
Short safety-related messages	Free format manually entered broadcasted to specific receiver or all vessels

Table 1: AIS data content and description, from [1].

2.1.1 Estimated Time of Arrival

Estimated time of arrival is the estimated time when a vessel will reach its destination, typically transmitted to the correct authorities 24 to 72 hours before arrival [3, 4]. The means of communicating this ETA varies and the accuracy of the ETA is not guaranteed to be within any margin of error. Ports, being very complex infrastructures with many moving parts, has to operate with certain uncertainties of

vessels arrival times and required demands and it is essential to plan port operations for a 24-hour period at a minimum [5]. All uncertainties that can delay or alter any operations at port are therefore a risk.

2.2 HELCOM dataset

Introduction to HELCOM

Describe the features and quirks with the HELCOM dataset as it is a processed dataset that contains AIS data already processed from the raw format. Merging of A and B AIS message types

HELCOM dataset has no ETA entries meaning to use this dataset for ETA prediction the ATA is calculated for all points on the route until the vessel has arrived, the true arrival time. Can't compare vessels ETA with the estimation but knows the estimated time left

Showing examples of the data as plots of the Baltic, also shows the area covered

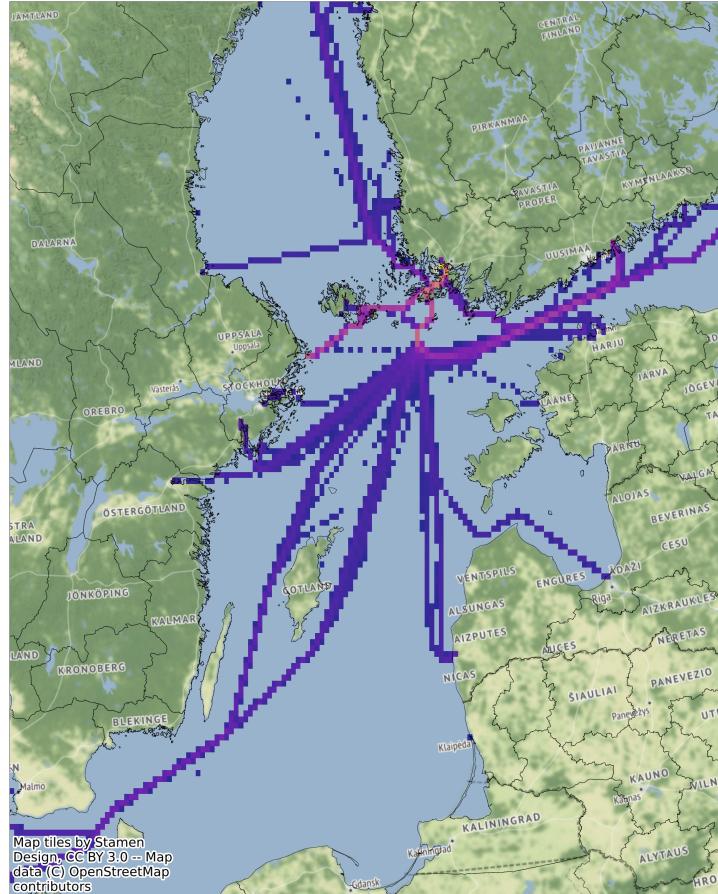


Figure 2: Heatmap the Baltic Sea with a limit of more than 10 unique messages per grid cell.

2.2.1 Description of HELCOM dataset features

The HELCOM dataset have twelve unique features for each entry described in Table 2. There are bad rows in the dataset, i.e. rows that have a or many bad values.

Name	Description	Value
timestamp	Unix epoch time in milliseconds when AIS message was created	Min 1230768000000
mmsi	Maritime Mobile Service Identities, unique for each vessel can change	9 digits long
lat	Latitude position when AIS message was generated	Coordinate in WGS 84
long	Longitude position when AIS message was generated	Coordinate in WGS 84
sog	Speed over ground in knots	0.1 knot resolution
cog	Course over ground in degrees relative to true north	0.1 degrees
draught	Vertical distance from waterline to the keel	0.1 meters
dimBow	Reference point for position of positioning system on the vessel	Meters from bow
dimPort	Reference point for position of positioning system on the vessel	Meters from port side
dimStarboard	Reference point for position of positioning system on the vessel	Meters from starboard side
dimStern	Reference point for position of positioning system on the vessel	Meters from stern
imo	Unique identifier for each vessel, does not change	7 digit identifier

Table 2: Variables in HELCOM data set and description.

2.2.2 Statistical information

Unevenness of the recorded messages (time intervals distribution even) but large gaps in routes. Reasons can be merging of many databases, preprocessing of the data to unify the records. The number of "good" valid routes for training is much lower than the total number of routes. Some of that could be handled by generating routes (data) where there are gaps. In a realistic dataset with AIS data there are going to be faulty data and missing, so HELCOM data is quite accurate to real data without generating missing data.

Vessels physical features and the grouping of different vessels by size.

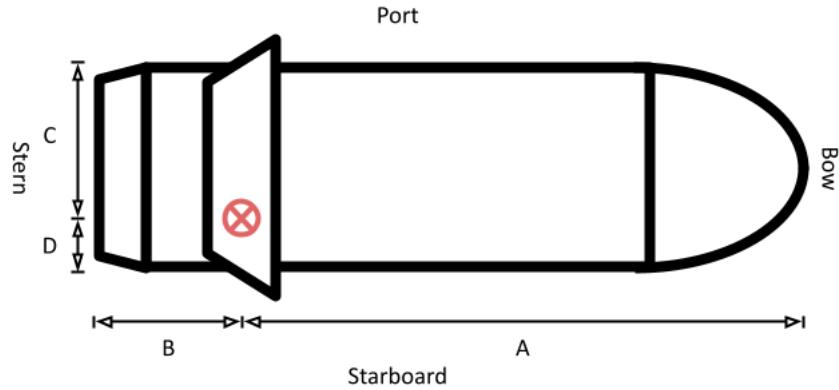


Figure 3: Vessel class definition from HELCOM data, the positioning system marked with red. A is *dimStern*, B is *dimBow*, C is *dimPort* and D is *dimStarboard*.

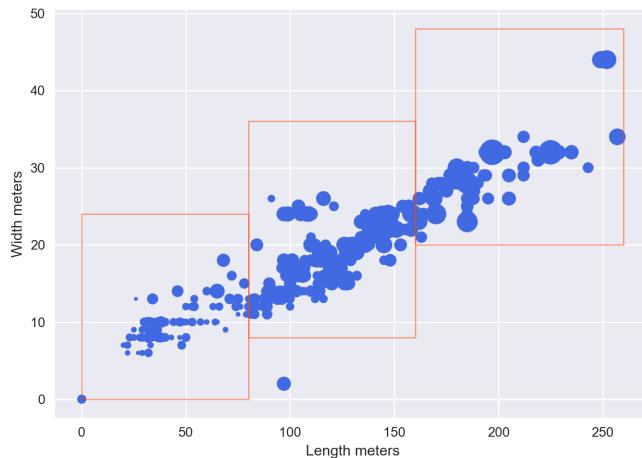


Figure 4: Defined vessel classes and distribution of all unique vessels for a 5 year period.

2.3 Port of Naantali

The amount of data collected travelling to FINLI from the HELCOM dataset, rather busy port approx. 2 % of the total data per month

The projects definition of the port, the bounding box with image.

3 Recurrent Neural Networks

Introduction to what neural networks and more precisely recurrent neural networks and why they are ideal for solving this problem (remembering the order of inputs and time is linear which means all previous timesteps have an impact on the next)

3.1 Long Short Term Memory

Description of what exactly LSTM networks are

4 Implementation

Python

4.1 Libraries used

Keras

Tensorflow (GPU)

Pandas

Numpy

Geopandas

Packages and programming language used.

4.2 Data pre processing

What range of data was used for the final model and reason for so. The problems with AIS data and getting clean routes from start to finish and primarily enough routes for training and testing.

Processing the HELCOM data to extract routes that can be used to train a model with the goal in mind.

4.2.1 Algorithm for extracting routes from dataset

Explain the algorithm behind getting the routes from the raw data set and that the idea can be applied to any port or area of interest really by defining the area of interest.

The largest gaps allowed

The port chosen for the evaluation has its caveats and should perhaps be identified.

Validation of the routes extracted i.e gaps, no shorter than n number of messages where the vessel is going, not returning to port, longer than n hours, no faulty data etc.

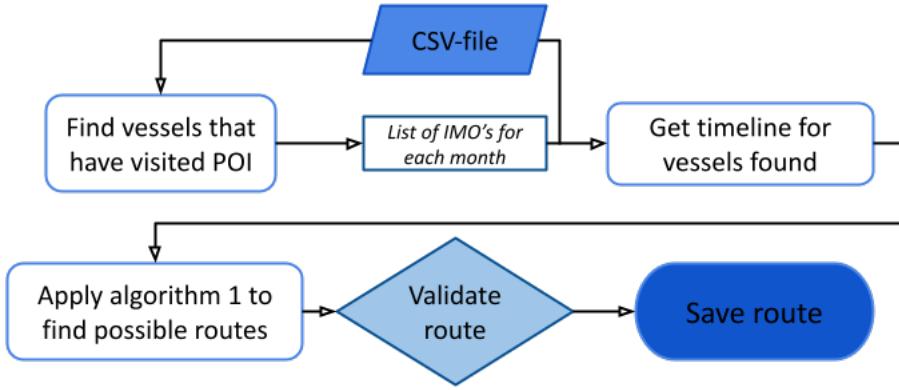


Figure 5: Getting routes.

POI (port of interest could also be thought of as Area of Interest). Timeline is the vessels historical data which is fed into algorithm 1. Route validation according to rules and then save the complete route.

4.2.2 Coordinate accuracy

Scaling the accuracy from raw data to a suitable accuracy. GIS decimal degrees

Vessel travelling at 14 knots (25 km/h) and a message interval of 10 minutes approx. will move 2.2 nmi (4.1 km) per message

4.2.3 Feature selection

The features chosen for the final training

Latitude, longitude, sog, cog, vessel class, draught, ?distance?

The target is Time To Destination, how many minutes are left to the destination

4.2.4 Time series data preparation and cleanliness

Further details on how the processed routes are handled to generate the training data to utilize multiple timesteps per prediction which improves performance. Also why the number of timesteps per prediction was chosen, with the time normalized data and how many steps then per time window.

The largest allowed difference between messages in the time normalized data

4.3 ML model

Description of the neural network model used and tested to find the optimal performer

4.4 Comparison model

!!! If the travel distance left to destination is used in nmi for example, test the accuracy against simply calculating time left by the current speed and distance left

5 Results

5.1 ETA prediction

5.2 Navigation in the archipelago

Models difficulties to predict ETA within the archipelago

6 Discussion

Discussion about the results and validity future work

7 Conclusion

The possibility to use historical AIS data, example HELCOM dataset, to train a model on predicting the ETA for ports that have good coverage of vessels inbound or some amount of routes coming to the port

References

- [1] Internation Maritime Organization, *Revised guidelines for the onboard operational use of shipborne automatic identification system*, 2015. [Online]. Available: <https://www.imo.org/en/OurWork/Safety/Pages/AIS.aspx> (visited on 12/14/2021).
- [2] “Commission Directive 2011/15/EU of 23 February 2011 amending Directive 2002/59/EC of the European Parliament and of the Council establishing a Community vessel traffic monitoring and information system,” 2011. [Online]. Available: <https://eur-lex.europa.eu/eli/dir/2011/15/oj>.
- [3] A. Veenstra and R. Harmelink, “On the quality of ship arrival predictions,” *Maritime Economics & Logistics*, vol. 23, no. 4, 655–673, 2021. DOI: 10.1057/s41278-021-00187-6.
- [4] “Directive 2009/16/EC of the European Parliament and of the Council of 23 April 2009 on port State control,” 2009. [Online]. Available: <http://data.europa.eu/eli/dir/2009/16/oj>.
- [5] G. Fancello, P. Claudia, M. Pisano, P. Serra, P. Zuddas, and P. Fadda, “Prediction of arrival times and human resources allocation for container terminal,” *Maritime Economics & Logistics*, vol. 13, pp. 142–173, 2011. DOI: 10.1057/mel.2011.3.