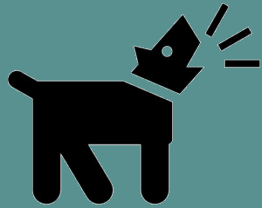


Data Reduction Techniques applied on Automatic Identification System Data



FAB LAB

Claudia Ifrim*, Iulian Iuga**, Florin Pop*, Manolis Wallace***, Vassilis Pouloupoulos***

* - University Politehnica of Bucharest, Bucharest, Romania

** - Independent Researcher, Bucharest, Romania

*** - Knowledge and Uncertainty Research Laboratory, University of the Peloponnese, Tripolis, Greece



Outline

- Scope and motivation
- What is an Automatic Identification System (AIS)?
- Analyzed data
- AIS Data Pre-Processing
- AIS Data Reduction
- Results
- Conclusions
- Future work



Scope and motivation

In recent years, the constant increase of waterway traffic generates a high volume of Automatic Identification System data that require a big effort to be processed and analyzed in near real-time. In this paper, we analyze an Automatic Identification System data set and we propose a data reduction technique that can be applied on Automatic Identification System data without losing any important information in order to reduce it to a manageable size data set that can be further used for analysis or can be easily used for Automatic Identification System data visualization applications.



What is an Automatic Identification System (AIS)?

The Automatic Identification System (AIS) is an automated tracking system used on ships and by vessel traffic services (VTS) that broadcasts in an interval of seconds information, such as unique identification of the ship, position, course, speed and navigation status, to other nearby ships, AIS base stations and satellites [1].



Analyzed data

- the AIS data set that we used for our experiments contains information retrieved from the area of the Black Sea and includes a number of 136,008,000 records;
- a PostgreSQL database with PostGIS extension is used to store the information within the messages. The information recorded was the decoded information and thus the two tables include:
 - static information - includes all the data that are related to the physical information of a vessel (type of vessel, length, year of construction, etc.)
 - dynamic information - includes latitude, longitude, speed, etc.
- we analyze the dynamic data;



AIS Data Pre-Processing

- Some information is eliminated as an initial cleanup procedure. This includes searching for the following malformed data and removing them:
 - Coordinates greater than 180, -180 latitude and 90,-90 longitude
 - The 0,0 location.
- This procedure removes almost 20% of the records in the dataset.



AIS Data Reduction

Analyzing the messages transmitted by a single vessel on a specific voyage, we can observe that the only attributes that are constantly changing are the ship location and the timestamp. We also observe that after a period on time attributes like speed and heading are also changing. Based on our observations on the AIS data set we conclude that the attributes **location, speed, heading and timestamp** can be used to develop our reduction algorithm.



AIS Data Reduction

- we will extract all the unique MMSI (Maritime Mobile Service Identity) values;
- for every unique MMSI value we will extract all its records in chronological order;
- the first record is considered a relevant record and its values for attributes like long, lat, speed, heading and timestamp are used as base values for further comparisons;
- iterating through all the records of the MMSI we will compare the selected attributes values



AIS Data Reduction

- if the values of lat, long and timestamp are equal the record is considered duplicate and is marked as unimportant
- if the values are different we will compare the speed and heading (if those values are higher than our tolerance values compared to the base attribute values, then the record will be considered important)
- the values used for further comparisons will be updated with the ones of the latest record marked as important.



Results

- our initial dataset contained 136 008 000 records (area Constanta port, Romania);
- we removed incorrect records in an initial cleanup and reduced the dataset with aprox. 20%;
- for this dataset we followed the algorithm described using different parameters for speed and heading of the vessels.



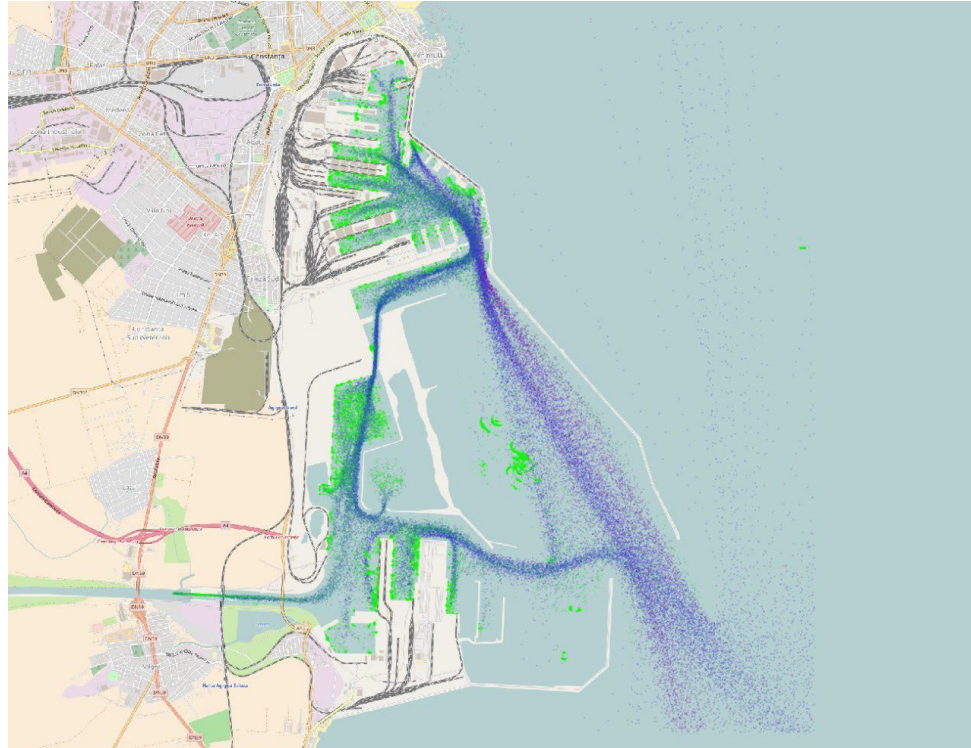
Results

Initial no. of records	Unique MMSI	Speed difference param	No. of records after reduction
752 552	458	less than 0.1 knots	248 743
752 552	458	less than 0.15 knots	204 338
752 552	458	less than 0.2 knots	202 248
752 552	458	less than 0.5 knots	187 881
752 552	458	less than 1 knots	177 845

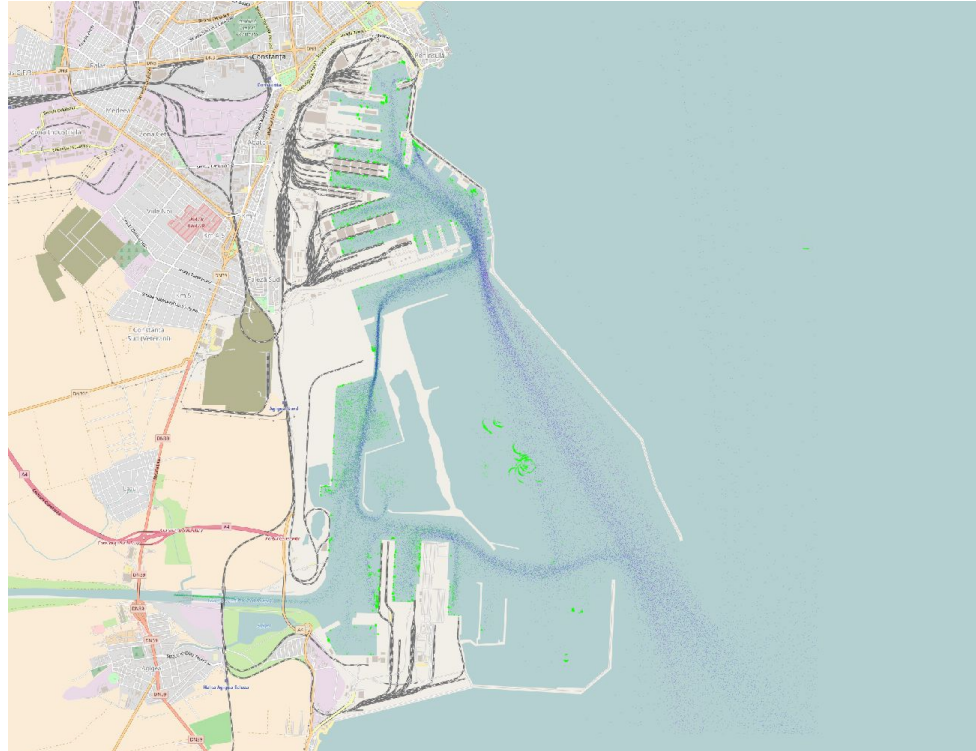
Results - Density map for initial dataset



Results - Density map for reduced set (speed difference less than 0.2 knots)



Results - Density map for reduced set (speed difference less than 0.5 knots)





Conclusions

As a conclusion for our experiment we consider that our reduction algorithm can be successfully used on AIS datasets (we preserve unaltered information for speed, heading, position and path of vessels) and the reduced information can be easily managed by applications that can be used in ports for the organization and planning of maritime traffic especially within ports or other dense traffic areas.



Future work

- create a real-time service for analyzing in real time the data produced by AIS;
- provide a near real-time API that will be able to reduce the volume of AIS data;
- adjust the parameters of the algorithm in order to achieve more efficient levels of reduction.



References

1. What is the Automatic Identification System (AIS)?
(<https://help.marinetraffic.com/hc/en-us/articles/204581828-What-is-the-Automatic-Identification-System-AIS->)
2. C++ decoder for Automatic Identification System for tracking ships and decoding maritime information (<https://github.com/schwehr/libais>)



thank you!

ifrim.claudia@gmail.com



Q&A

