HACKtheMACHINE SEATTLE

TRACK 2: Data Science and the Seven Seas: Collision Avoidance

International regulations for preventing collisions at sea (COLREGS) codify long-established international norms for maritime navigation. Mariners are held accountable to the principles in COLREGS for safely handling their vessels. However, many situations that mariners frequently encounter, such as three vessels converging, are not covered by COLREGS. Similar to automotive traffic laws, the safest maneuver, in a given situation, depends on many factors. For example, turning right may be legal in two different situations. At an intersection on a country road, this maneuver would be safe and failing to turn would result in an angry train of drivers behind you. If you tried that same right turn in Times Square in NYC you would find yourself trapped in a swarm of pedestrians, unable to safely make the turn. The context-dependent nature of COLREGS makes it particularly difficult for autonomous systems, which lack human judgement. This challenge will use data from ships underway on the high seas to develop algorithms to assist the Navy with preventing collisions for human-operated and autonomous vessels.

AIS Data:

For this challenge, your primary data is collected from the Automatic Identification System (AIS). Ships carry AIS to provide unique identification, position, course, and speed data to other ships and shore stations. AIS is intended to assist the crew with safe navigation and allow maritime authorities to monitor vessel movements. The International Maritime Organization requires AIS to be fitted aboard international voyaging ships with 300 or more gross tonnage (GT), and all passenger ships regardless of size.

AIS data is enormous and rich (the 2016 United States dataset alone has 9.4 billion records), but it was designed as a safety feature for local collision avoidance communication between ships at sea, not to be analyzed holistically. As such AIS data provides a huge opportunity for understanding the contextual nature of COLREGS encounters for ships at sea. It also presents significant challenges including:

- **Noise:** AIS systems transmit radio signals in the Maritime VHF band, which is inherently noisy.
- **Transmission Gaps:** There are often jumps in the data due to spotty AIS signals, low density of receivers, or vessel operators turning off the signal to avoid detection.
- **Congestion:** In crowded areas, bandwidths become congested and signals interfere with one another.
- **Errors:** Sometimes ships appear in impossible places like on land or airports or vessel operators mis-enter AIS codes. This might be a mistake or malfunction or can be indicative of malicious activity (https://www.wired.co.uk/article/black-sea-ship-hacking-russia).

Determining “normal” behavior is incredibly difficult and something you can help the US Navy tackle through the following 2 challenges.

Challenge 1: Identifying Interactions between Ships

Out of the millions of data points in the AIS dataset can you find the most efficient algorithm to identify potential collisions?

COLREGs interactions are defined by situations where there are constant bearing and decreasing range between two ships. In other words, two ships are headed for collision unless one or both change course or speed. COLREGs also assume that two ships are visually in sight of one another. You can assume they are a maximum of 4 nautical miles away. This challenge is about developing a screening tool for ships interactions within an AIS dataset, which is akin to finding the proverbial needle in the haystack.

As a starting point, consider looking for places where ships operate with a higher than average risk of collision. Ships frequently come close to each other in high traffic areas, like port entrances, but often the ships are not moving, or the interactions are governed by prescriptive criteria, like traffic lanes. The more interesting interactions are in areas where the traffic patterns have not explicitly defined.

You may use any of the AIS data to address this challenge, but we would suggest looking at a high traffic area with rules, like the Los Angeles/San Diego area (UTM 11) and an area that has less rules and more open ocean, like The Caribbean (UTM 17 and 18, less than 27 degrees latitude), as a starting point we have pre-loaded some data into the ESRI platform and the GitHub repository.

The judging criteria for the challenge will include:

Up to 50 points for developing a computationally efficient screening tool. We need a way to go from millions of data point to the few hundred thousand that might represent interesting interactions between ships. There are a couple of examples of approaches in the GitHub repository, but none are efficient enough to handle the size and scale of the full AIS dataset.

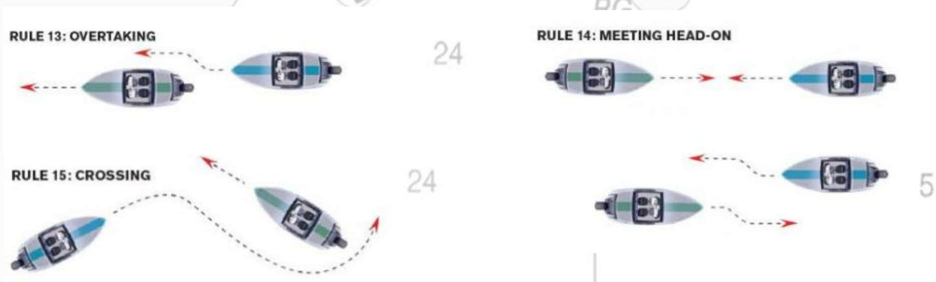
Up to 40 points for techniques to refine the initial screening tool. Once we have a screening tool, we need to go from hundreds of thousands of candidates to those that are actually COLREGs interactions. The candidate interactions might include activities like towing or ships in port that need to be filtered out. In some cases, the 1-minute intervals in the Marinecadstre dataset will not be sufficient to capture the nuances of the interaction and identification will require linking to another dataset, like Spire, to get the ships' tracks. We prefiltered the Jan 2106 UTM 11 (LA/San Diego) dataset by proximity (4 nautical miles) if you want to use that as a starting point. From there you can refine your algorithms in any way you see fit.

Up to 10 points for data presentation and visualization.

Challenge 2: Behavioral Models for Ships' Interactions

Having accurate models of the real-world behavior for ships is a valuable tool for preventing collisions for both manned and autonomous vessels. This challenge uses the AIS data to find patterns in vessel behavior by connecting the observed behavior to variables, such as vessel type, location, or features found on Nautical Charts. By determining which variables are linked to vessel behavior we can begin to develop probabilistic models to prevent collisions.

We will use COLREGS interactions to describe the interactions between ships. First, there are 3 types of encounters: overtaking, crossing, or meeting.



Images Adapted from article originally published in Bassmaster Magazine, June 2012

Second, COLREGs apply different standards to vessels in restricted waters such as narrow channels, and traffic separation schemes, and large vessels that are restricted in their ability to maneuver. Understanding this context requires linking the AIS data to Electronic Navigational Chart data.

Third, vessels have different hierarchies depending on the situation. Powered vessels must give way to sailing vessels, which must give way to vessels engaged in fishing, except under certain circumstances. Modeling ships behavior requires knowing the type of vessel and what it is doing during the encounter. There are several good websites for discovering vessel details including https://www.marinetraffic.com/. The ship's activity can be inferred from AIS data (as VesselType) or observing the ships behavior.

The USCG COLREGs Quick Reference as well as a compete COLREGs may be helpful as references.

We have a sample dataset that includes one interaction of each of the COLREGs encounters: crossing, overtaking, and meeting. The data is available on the ESRI platform and the GitHub repository. You will get extra points for using the results of challenge 1 as your inputs into challenge 2.

The judging criteria for the challenge will include:

Up to 30 points for model completeness and applicability. Developing techniques for filling in missing AIS data, adding variables, or predicting the behavior of vessels.

Up to 20 points for presentation and visualization.

Data Description

There are several AIS sources for this challenge.

1) The MarineCadastre data (https://marinecadastre.gov/ais/) contains AIS data from land-based USCG receivers in U.S. coastal waters. Records are filtered to one minute and formatted in zipped, monthly files by Universal Transverse Mercator (UTM) zone. The data from 2015-2017 is available in comma separated variable (CSV) format and contains the fields below.
2) Our sponsor, Spire, has provided combined Satellite and terrestrial AIS that has more frequent transmissions from ships and isn't limited by ships having to be close to land-based receivers. The Spire data is pre-loaded into the ESRI. The Spire data is divided into static and position messages, so to get the full complement of fields shown below (plus a few additional fields) you will have to join the data on MMSI.

AIS Data Fields

MMSI	A series of nine digits uniquely identifying ship stations
BaseDateTime	Timestamp – Coordinated Universal Time (UTC) time accurate to nearest second when this data was generated
LAT	Latitude – to 1/10,000 minute
LON	Longitude – to 1/10,000 minute
SOG	Speed over ground – 0 to 102 knots with 0.1 knot resolution
COG	Course over ground – relative to true north to 0.1 degree
Heading	True Heading – 0 to 359 degrees from gyro compass
VesselName	20 characters to represent the name of the vessel
IMO	IMO (International Maritime Organization) ship identification number – a seven-digit number that remains unchanged upon transfer of the ship's registration to another Country
CallSign	International radio call sign – up to seven characters, assigned to the vessel by its Country of registry
VesselType	Type of ship/cargo
Status	Navigation status – "at anchor", "under way using engine(s)", or "not under command"
Length	Dimensions of ship – to nearest meter
Width	Dimensions of ship – to nearest meter
Draft	Draught of ship – 0.1 meter to 25.5 meters
Cargo	AIS Vessel Codes

Challenge Logistics

Everything you need to get started can be found on GitHub. The GitHub repository includes data, sample code, and submission instructions and will be made public shortly before the competition begins.
https://github.com/FATHOM5/Seattle_Track_2

Everything you need for geospatial analysis has been provide by HACKtheMACHINE sponsor, ESRI. ESRI has setup an enterprise portal for geospatial analysis. A large amount of the data has been preloaded into the ESRI environment along with a number of geospatial analysis tools. Each team will be given credentials to access the environment prior to the competition.
https://htm.esri.com/portal/home/index.html. Tutorials for the ESRI tools can be found in the participant resources section at the bottom of the https://www.hackthemachine.ai/track2/ webpage.

You may use any methods and tools for the challenge. As part of the judging process you will be submitting your code and datasets and giving a short presentation (3 minutes) to explain your techniques and results. We will have both technical and subject matter mentors available

FRIDAY SEPT 21st
1200: TUTORIAL: ArcGIS Enterprise Overview
1300: TUTORIAL: ArcGIS Pro Introduction
1400: TUTORIAL: ArcGIS and Machine Learning
1500: TUTORIAL: Analytics in AWS
1800: Kickoff and Reception at Pacific Science Center

SATURDAY SEPT 22nd
0900: Event Kickoff
0930: Break into Tracks: Intro / Agenda / Tech Brief
0945: Team Formation
1000: Hacking Begins
1300: Mentor Office Hours
1515: Executive Pitches
2200: Building Closes

SUNDAY SEPT 23rd
0730: Building Opens
1000: Review and Pitch Mentoring (Required)
1100: Working Session
1300: Presentation Prep + Elimination Round
1400: Presentation to Track Judges
1500: Awards