A vessel risk framework to identify vessels involved in illegal, unregulated, and unreported fishing

Technical Report

Iván Higuera-Mendieta Shubham Tomar William Grimes

August 25, 2017

Abstract

Data science is increasingly being applied to optimise, or improve operations in business and industry. There is also a movement gaining popularity to apply data science approaches to solve problems relating to social good, for example in environmental, healthcare, or government services. A framework is here outlined that demonstrates an approach to improve detection of illegal, unregulated, and unreported (IUU) fishing using multiple data sources. Automatic identification system (AIS) data was used to predict whether a vessel is fishing, and generate components to indicate the likelihood of a vessel engaging in IUU behaviours. These AIS data were then correlated with satellite imagery and vessel images extracted. This document describes the data sources relevant to this problem, the approach taken, as well as evaluation of our methods, and future work.

Contents

1 Introduction

Every year around 26 million tonnes of seafood worth close to $24 billion are extracted from the planet's oceans by illegal, unreported and unregulated (IUU) fishing. These IUU fishing techniques include: extracting fish from waters of other nations or designated marine protected areas, catching fish using illegal and ecologically damaging techniques, or under-declaring fish by transshipment at sea. Such practices are especially rife in the rich tropical waters of Southeast Asia due to the burgeoning demand in the region, and challenges of enforcement. This project sets out to address the problem of IUU fishing by correlating multiple data sources, and using data science techniques to identify vessels involved in IUU fishing.

Overfishing and IUU fishing have lead to huge declines in fish stock and some species such as tuna have declined by over 90%. Our project partner the World Economic Forum (WEF) is committed to improving the state of the world by engaging business, political, academic, and other leaders of society to shape global, regional, and industry agendas. In particular the WEF promotes initiatives to improve ocean governance, food chain sustainability, and environmental conservation.

In partnership with the WEF this DSSG Europe project has set out to create an open-source tool combining multiple data sources to help combat IUU fishing. This proof-of-principle study of fishing in the Torres Strait aimed to demonstrate how data aggregation from sources such as satellite imagery, synthetic aperture radar (SAR) and automatic identification systems (AIS) can be correlated and used with data science methods such as object recognition and anomaly detection to aid in identification of illegal fishing. This data science approach to detecting IUU fishing could ultimately improve enforcement, guide governance, and inform policy decision making.

There are many challenges associated with detecting IUU fishing. Firstly, the world's ocean comprise the majority of the Earth's surface (71%). This is a large area to inspect and hence the volumes of data involved are large. There are many vessels in this large area relating to commercial, leisure, or fishing activities. Systems to detect and track these vessels such as automatic identification systems (AIS) tend to be implemented nationally in vessel management systems (VMS), hence there is little standardization of the data format. Satellite data on the other hand is relatively infrequent, can be obscured by cloud cover, and may not have good coverage in the ocean. Combining these data sources can be difficult to find appropriate images and AIS data. There is also a distinct lack of high-resolution open-source data sources in this domain, due to the cost of data collection.

A socially desirable outcome for this project would be to successfully demonstrate how these data can be used to identify IUU fishing. In partnership with the WEF and organisations this can be conveyed to policy and decision makers to expand the study. A socially desirable outcome would be to improve detection of illegal fishing via these data sources, a secondary outcome of this would be to provide improved enforcement of illegal fishing, which in turn would improve regulation as it becomes harder to evade capture. The result of this would be to promote more sustainable fishing practice and environmental conservation.

2 Previous work

Ship detection is a relevant enforcement tool, and also can serve as a way understand a myriad of behaviors in the ocean, such as piracy, illegal maritime trafic, and UII fishing. A complete set of literature have used different ship detection algorithms and different sources of geospatial data. For instance, Tello, Lopez-Martinez, and Mallorqui (2005) uses satellite-based synthetic aperture radar (SAR) and a discrete wavelet transform to capture spectral signals of vessels in the ocean. The use of SAR data to this task is common in the literature (Margarit, Milanés, & Tabasco, 2009; Brusch et al., 2011; Corbane, Marre, & Petit, 2008; Paes, Lorenzzetti, & Gherardi, 2010). [1] The ability of the SAR sensors to capture signals even in the presence of higher percentage of cloud coverage, give radar data an advantage over visible (optical) data. Nonetheless, vessel detection with SAR data has limitations for identifying small ships, limited coverage, and its automatic interpretation can be a cumbersome process (Zhang, Wu, Zhang, Huang, & Tian, 2006).

[1] Most of these documents rely on the *TerraSAT* proyect data which have a resolution up to ≈ 16 meters.

Corbane, Najman, Pecoul, Demagistri, and Petit (2010), shows a first apporach to vessel detection using panchromatic high-resolution imagery from the SPOT-5 program. The authors used a wavelet transform, which decompose the image according to light frecquency profiles and used a novel pre-processing approach that allows having rapid classifications compared with other algorithms. Lebona, Kleynhans, Celik, and Mdakane (2016) also uses optical data to track vessels, although the authors use the NASA-VIIRS nightlight data. The results of the classification algortihm are cross-validated using vessel positional AIS data, confirming the correct indetification of 5000-6000 ships per night, but is not clear if this approach is suitable to identify small targets, like fishing vessels.[2].

Image classification is not the only approach to track vessels on the ocean. Positional messages are also used to classify ship activity

3 Data sources

3.1 Automatic identification system (AIS)

Vessel location for the period from May 2016 to June 2017 was captured using the Automatic Identification System (AIS) transponders installed in every vessel with a length greater or equal to 30 meters. These transponders broadcast two types of messages. First, they transmit positional signals every 10 seconds, reporting GPS coordinates and navigational features of the vessel (i.e., speed, course, and heading). The second type of message is static and reports constant features of the ship, such as name, callsign, length, and destination.

Beacons near the shores capture these messages, working as an avoiding collision system with the sea coast and other vessels. This makes AIS an important component of maritime security (Tetreault, 2005). Coastal receivers do not always capture AIS signals away from shore.[3] Nonetheless, we use satellite captured AIS signal provided by Spire; this allows us to retrieve vessel tracking data from all over the world, and not only from vessels near seashores.Some caveats arise with the use of this source. First, AIS signals fetched by satellites can *bunch together* bunch together leading to data loss, since the satellite is only able to process a limited number of signals at a time. Second, accuracy is not always certain. Hence there is a possible measurement error in the location and other features of the vessels. Lastly, there is the possibility of AIS tampering in vessels involved in illegal activities, such as UUI.

3.2 Satellite imagery

To account for a visual validation of the AIS data, we used different sources of proprietary satellite imagery for the same timeframe of the positional AIS data. We rely on two sources of high-definition satellite imagery. The first source is Digital Globe (henceforth GBDX) which reports images with visual and Near Infrared (NIR) bands with a resolution of $\approx$ 3 meters per pixel. Second, we used imagery from Planet, which has higher revisit times that GBDX, but at the cost of lower resolution (from $\approx$ 10 m. to $\approx$ 3 m. over the equator).

Images from both sources were orthorectified and projected into a common grid. The first process corrects possible terrain distortions due to higher elevation angles.[4] The second process projects the images into a common grid to georeference them into a common grid, usually under a *UTM* local projection. Despite the availability of more advanced image processing tasks, as *Pansharpening* or *Atmospheric compensation*, we decided to use a single approach. Also, some advanced image processing tasks merge visual and NIR bands, hindering the use of a multispectral or thermal analysis.

[2]VIIRS day-night band, as its predecesor, the NOAA-OLS, has a lower resolution (1 km^2 at the equator), and other problems like overglooming in the coastlines that can yield false positives. To read a comprehensive assessment of the use of nightlight data to data analysis see Min (2015)

[3]According to Spire, the coastal receivers can only capture data in an 80 kilometers range.

[4]This angle is commonly known as Nadir. The terrain displacement occurs when satellite sensors are not orthogonal to the earth terrain. Thus the optimal Nadir angle is 90°.

4 Detecting weather a vessel is fishing at a given point

One of our primary objectives is to identify vessels that are fishing in maritime protected areas.

International Union for Conservation of Nature - IUCN has defined MPAs as follows: 'A clearly defined geographical space, recognised, dedicated and managed, through legal or other effective means, to achieve the long-term conservation of nature with associated ecosystem services and cultural values'.

MPAs are classified into several categories based on the fishing restrictions. Restrictions may vary from protecting particular species to no fishing allowed.

Being able to identify weather a vessel is fishing at any point along its trajectory will help us identify vessels fishing in MPAs and Exclusive Economic Zones.

4.1 Data

We are using the AIS data provided by Spire. AIS transceivers send positional data every 2-10 seconds and static data every 6 minutes.

The positional AIS data contains the following fields:

- The vessel's Maritime Mobile Service Identity (MMSI) âĂŞ a unique nine digit identification number.
- Navigation status - "at anchor", "under way using engine(s)", "not under command", etc.
- Rate of turn âĂŞ right or left, from 0 to 720 degrees per minute
- Speed over ground âĂŞ 0.1-knot (0.19 km/h) resolution from 0 to 102 knots (189 km/h)
- Positional accuracy: Longitude âĂŞ to 0.0001 minutes Latitude âĂŞ to 0.0001 minutes
- Course over ground âĂŞ relative to true north to 0.1Âř
- True heading âĂŞ 0 to 359 degrees (for example from a gyro compass)
- True bearing at own position. 0 to 359 degrees UTC Seconds âĂŞ The seconds field of the UTC time when these data were generated.

The static data contains the following fields:

- IMO ship identification number âĂŞ a seven digit number that remains unchanged upon transfer of the ship's registration to another country
- Radio call sign âĂŞ international radio call sign, up to seven characters, assigned to the vessel by its country of registry
- Name âĂŞ 20 characters to represent the name of the vessel
- Type of ship/cargo
- Dimensions of ship âĂŞ to nearest meter
- Location of positioning system's (e.g., GPS) antenna on board the vessel - in meters aft of bow and meters port or starboard
- Type of positioning system âĂŞ such as GPS, DGPS or LORAN-C.
- Draught of ship âĂŞ 0.1 meter to 25.5 meters
- Destination âĂŞ max. 20 characters
- ETA (estimated time of arrival) at destination âĂŞ UTC month/date hour:minute

We have access to an year of worldwide AIS messages. We also used the training data provided by Global Fishing Watch. The data contains positional data and the labels for weather the vessel was fishing at that point or not. This data set was labeled for 90 vessels over a period of 3 years.

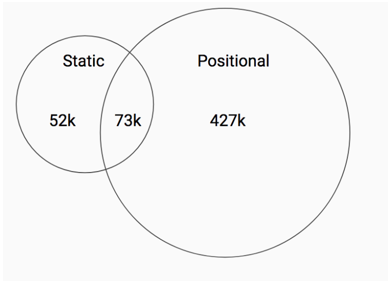


Figure 1: Number of positional and static AIS messages for a period of 1 year

5 Modelling

Here we describe an approach to use AIS data to identify at what points along their trajectory the vessels were fishing.

5.1 Feature generation

After discussions with domain experts, we came up with a list of features that might be good indicators of fishing behavior. We generated features like distance from shore, port, day or night during the time of navigation etc. We then fit a Random Forest model to compute how each of the feature contributes towards the prediction.

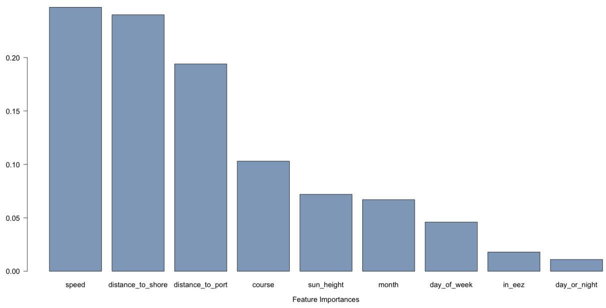


Figure 2: Feature importance for AIS positional features.

5.2 Training and cross-validation

Given we have a time series data, we split our data set into multiple temporal folds. For each fold, we split the train-test set as illustrated in the following figure.

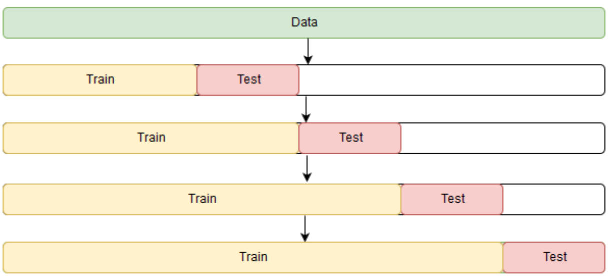


Figure 3: Temporal splits for each fold

We compute various metrics like precision, recall, false positive and true positives rates etc.

5.3 Model selection

After talking to our partners, we came to the conclusion that we want to optimize for precision of our models.

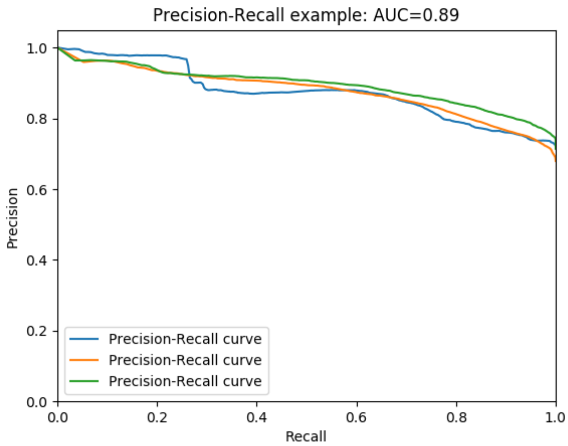


Figure 4: Precision-Recall curve for different models

6 Evaluation

In the absence of a labeled prediction set, we are employing a few heuristics to examine the performance of our models. For example, average fishing score for reported fishing vessels should in general, be greater than the same for vessels that don't report themselves as fishing.

Vessels spend most of their time being docked. Also, reported fishing vessels spend only a small fraction of their time actually fishing.

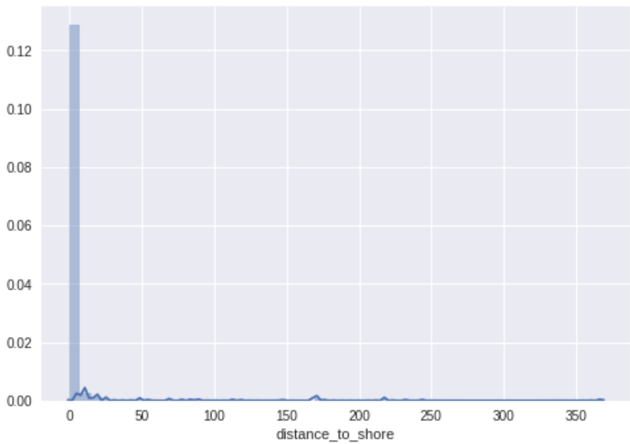


Figure 5: Distribution of distance to shore in the AIS messages

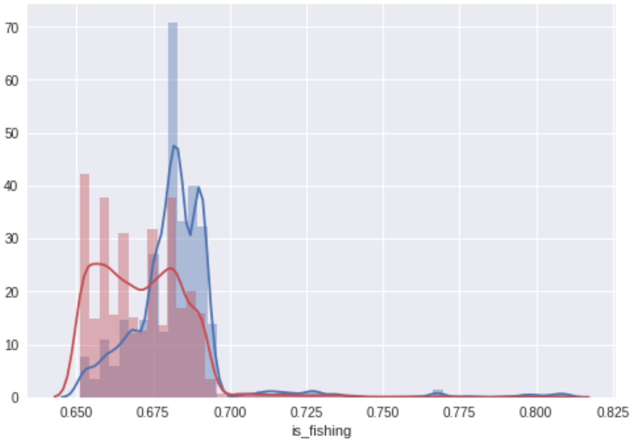


Figure 6: Fishing score distributions for reported and non reported fishing vessels.

6.1 Intersection AIS and Satellite imagery

Marine traffic detection using remote sensing approach is a commonly addressed task. Brusch et al. (2011) showed how AIS and Synthetic Aperture Radar (SAR) data could be used in conjunction to detect vessels in small portions where the satellite imagery tiles contain positional AIS messages. Corbane et al. (2008) follow a similar approach joining Vessel Monitoring System (VMS) and commercial satellite imagery to identify common features of shrimp boats. Although these endeavors are working solutions, there is still wanting a strategy that can bring different sources of data that can complement SAR data. One of these possible sources of data is high-resolution imagery, which until now has been restricted to commercial use only (Greidanus & Kourti, 2006).

Our approach follows the work previously outlined since it is combining positional GPS data with remote sensing data. Nonetheless, it departs from it since it is mainly focused on UII activities, and also uses worldwide satellite imagery and AIS positional data. This work, at least in this preliminary stage, can work to create a comprehensive image database that can serve ultimately to validate the AIS data accuracy, and also the feasibility of joining different data sources to understand risky behaviors related to UII.

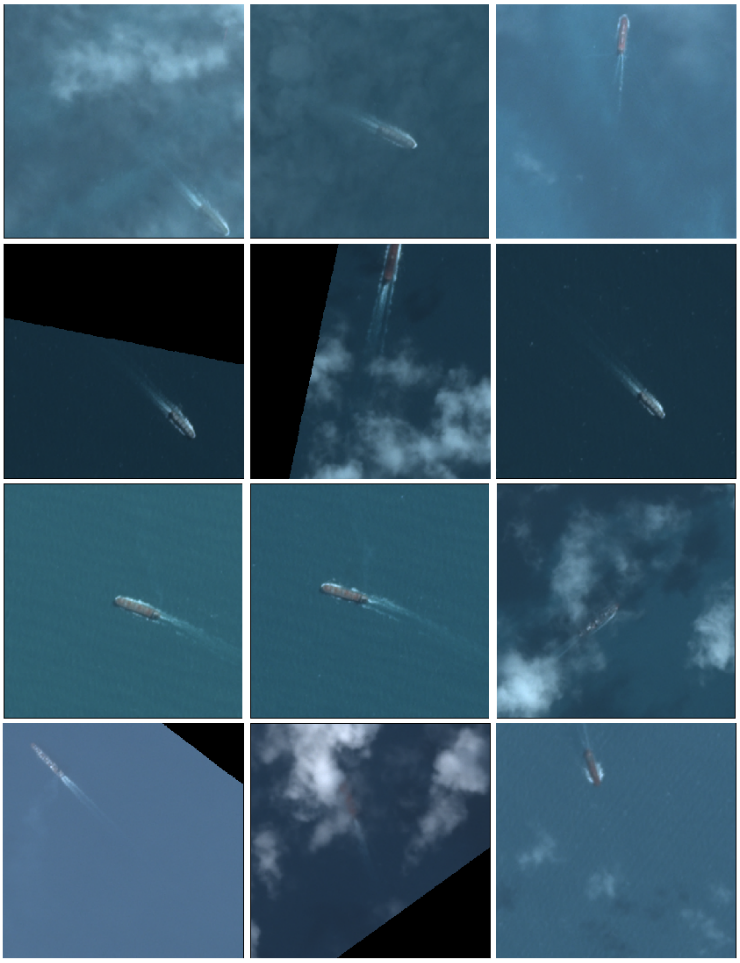
The process of image retrieval was a tree step operation. First, we retrieve all the available images in *Marine Areas* and *Ocean Areas* for our time frame (May 2016 to June 2017). The former are defined as areas is open waters; meanwhile the latter are the areas near the continental land. This process yielded 187.038 images.[5]. Second, we overlap the bounding boxes of the imagery tiles with the AIS positional signals. This process was not only a spatial join, but also a time join which took into account the difference of time between the picture capture timestamp and the positional AIS timestamp. Figure 7 shows the outcome of this step. In total, for Digital Globe, the intersection retrieved 1.072 AIS positional signals (487 unique vessels). Third, we crop the images using a 800 meters square buffer of each the AIS coordinates inside the overlap. The Figure 8 shows the final outcome of this process.

[5]Additional 18.476 images were retrieved using Planet imagery only for the Torres Strait.

Figure 7: Overlap Digital Globe and AIS data (May 2016 - June 2017)

Source: Digital Globe, Spire and OpenStreetMap.

Figure 8: Cropped images from Planet (May 2016 - June 2017)



Source: Planet and Spire.

7 Evaluation methodology

evaluate this

8 Future work

Lots will be done in the future

Acknowledgement

The authors would like to thank x, y, and z for their dedication.

References

Brusch, S., Lehner, S., Fritz, T., Soccorsi, M., Soloviev, A., & van Schie, B. (2011). Ship Surveillance With TerraSAR-X. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(3), 1092–1103. doi: 10.1109/TGRS.2010.2071879

Corbane, C., Marre, F., & Petit, M. (2008). Using SPOT-5 HRG Data in Panchromatic Mode for Operational Detection of Small Ships in Tropical Area. *Sensors*, *8*(5), 2959–2973. Retrieved from `http://www.mdpi.com/1424-8220/8/5/2959` doi: 10.3390/s8052959

Corbane, C., Najman, L., Pecoul, E., Demagistri, L., & Petit, M. (2010). A complete processing chain for ship detection using optical satellite imagery. *International Journal of Remote Sensing*, *31*(22), 5837–5854. Retrieved from `http://dx.doi.org/10.1080/01431161.2010.512310` doi: 10.1080/01431161.2010.512310

Greidanus, H., & Kourti, N. (2006). Findings of the DECLIMS projectâĂŤDetection and classification of marine traffic from space. *Proceedings of SEASAR 2006*, 23–26.

Lebona, B., Kleynhans, W., Celik, T., & Mdakane, L. (2016). Ship detection using VIIRS sensor specific data. In *2016 ieee international geoscience and remote sensing symposium (igarss)* (pp. 1245–1247). doi: 10.1109/IGARSS.2016.7729315

Margarit, G., Milanés, J. A. B., & Tabasco, A. (2009). Operational ship monitoring system based on Synthetic Aperture Radar processing. *Remote Sensing*, *1*(3), 375–392. doi: 10.3390/rs1030375

Min, B. (2015). *Power and vote. Elections and Electricity in the Developing World.* Cambridge: Cambridge University Press.

Paes, R. L., Lorenzzetti, J. A., & Gherardi, D. F. M. (2010). Ship Detection Using TerraSAR-X Images in the Campos Basin (Brazil). *IEEE Geoscience and Remote Sensing Letters*, *7*(3), 545–548. doi: 10.1109/LGRS.2010.2041322

Tello, M., Lopez-Martinez, C., & Mallorqui, J. J. (2005). A novel algorithm for ship detection in SAR imagery based on the wavelet transform. *IEEE Geoscience and Remote Sensing Letters*, *2*(2), 201–205. doi: 10.1109/LGRS.2005.845033

Tetreault, B. J. (2005, sep). Use of the Automatic Identification System (AIS) for maritime domain awareness (MDA). In *Proceedings of oceans 2005 mts/ieee* (pp. 1590–1594 Vol. 2). doi: 10.1109/OCEANS.2005.1639983

Zhang, F., Wu, B., Zhang, L., Huang, H., & Tian, Y. (2006). Illicit vessel identification in inland waters using SAR image. In *Society of photo-optical instrumentation engineers (spie) conference series* (Vol. 6419, pp. 64190S–64190S–8). doi: 10.1117/12.712973